BEUW

W.Q. LIB
ST. CLAIR R. (23)

STANDARDS DEVELOPMENT BRANCH OMOE
36936000009215

WATER CHEMISTRY DATA FOR THE ST. CLAIR RIVER AT CORUNNA, ONTARIO

as determined by continuous water quality monitor

Ministry
of the
Environment

Ontario

Copyright Provisions and Restrictions on Copying:

This Ontario Ministry of the Environment work is protected by Crown copyright (unless otherwise indicated), which is held by the Queen's Printer for Ontario. It may be reproduced for non-commercial purposes if credit is given and Crown copyright is acknowledged.

It may not be reproduced, in all or in part, for any commercial purpose except under a licence from the Queen's Printer for Ontario.

For information on reproducing Government of Ontario works, please contact ServiceOntario Publications at copyright@ontario.ca

WATER CHEMISTRY DATA
FOR THE
ST. CLAIR RIVER
AT CORUNNA, ONTARIO, AS
DETERMINED BY CONTINUOUS WATER QUALITY MONITOR

D.J. Poulton and M.D. Palmer
Water Quality Branch
Ontario Ministry of the
Environment
40 St. Clair Ave. W.
Toronto 195, Ontario (1973)

TABLE OF CONTENTS

Summary 1

1. Introduction 3

2. Description of Equipment 5

3. Statistical Analysis of Data 7

    3.1 Tape Translation 7

    3.2 Data Editing 8

    3.3 Trend Correcting 8

    3.4 Frequency and Time Distribution Analysis 9

    3.5 Time Series Analysis 10

    3.6 Markov Chain Analysis 12

4. Results and Discussion 15

    4.1 Data Preparation 15

    4.2 Frequency and Time Distribution Analysis 17

    4.3 Time Series Analysis 18

    4.4 Examples of Extreme Data Observed 22

    4.5 Markov Chain Analysis 24

Acknowledgements 26

References 27

APPENDICES

1. Description of the robot monitor and problems encountered during its operation 53

2. Comment on quality of various data records 59

3. Statistical analysis of frequency distributions 62

4. Operations and calibrations log 72

LIST OF TABLES

Table 1	Sample Robot Monitor Data	28
Table 2	Raw Data, April 1971	29
Table 3	Edited Data, April 1971	30
Table 4	Trend Corrected Data, April 1971	31
Table 5	Polynomial Fitting During Trend Correction	32
Table 6	Means and Standard Deviations of Corunna Robot Monitor Data	33
Table 7	Distribution Probabilities by Kolmogorov-Smirnov Test	34
Table 8	80% Confidence Spectral Maxima	35
Table 9	Some Analytical Data for St. Clair River Determined by Great Lakes Surveys	36
Table A2.1	Index of Data Tapes Obtained at Corunna	61
Table A3.1	Kolmogorov-Smirnov and Von Mises Test Results on Sampled Data	69
Table A3.2	Discrete Normal Distribution Tests	70

LIST OF FIGURES

Figure 1	Location of Robot Monitor	37
Figure 2	Raw Chloride, Frequency Distribution	38
Figure 3	Edited chloride, frequency distribution	39
Figure 4	Trend corrected chloride, frequency distribution	40
Figure 5	Autocorrelation coefficients for June 1971 chloride	41
Figure 6	Autocorrelation coefficients for June 1971 temperature	42
Figure 7	Autocorrelation coefficients for June 1971 dissolved oxygen	43
Figure 8	Autocorrelation coefficients for June 1971 conductivity	44
Figure 9	Variance density spectra for June 1971 chloride	45
Figure 10	Variance density spectra for June 1971 temperature	46
Figure 11	Variance density spectra for June 1971 dissolved oxygen	47
Figure 12	Variance density spectra for June 1971 conductivity	48
Figure 13	Dissolved oxygen frequency distributions for April-June 1971	49
Figure 14	State probability distributions for chloride, June 1971	50
Figure 15	State probability distributions for dissolved oxygen, June 1971	51
Figure 16	State probability distributions for conductivity, June 1971	52
Figure A1.1	Schematic drawing of robot monitor intake, August 1971	58

SUMMARY

The operation of a shore-based automatic digitally recording water quality monitor on the St. Clair River downstream from Sarnia during the period April 1971 - July 1972 is discussed. This monitor measures chloride, temperature, dissolved oxygen, pH and conductivity. Instruments of this nature employed singly are limited to detecting extreme values of chemical parameters which may be missed during routine sampling, as well as interparameter correlations and periodic variations. As data are collected at a fixed point some other checking is required to determine if the point is representative of the river quality. The data translation, conditioning and analysis requirements are extremely demanding on both personnel and computer time for such limited information. Operational problems with the intake and shore mounted equipment, particularly the mechanical components caused numerous breakdowns or resulted in unusuable data.

Typical ranges of data obtained were as follows: chloride, 20-40 mg/l; dissolved oxygen, 70-110% saturation; pH, 7.5-9.0 units; and conductivity, 200-350 μmhos/cm.

High values of chloride data collected during April 1971 correlated with meteorological data, and strongly suggested runoff of salt from city streets. In April-June 1971, the results obtained were characterized by generally non-normal frequency distributions and diurnal variations of

temperature, dissolved oxygen and pH consistent with solar heating and photosynthesis. In November 1971-January 1972, higher probabilities of normal distribution accompanied by the almost complete absence of periodic variations indicated the interaction of a large number of phenomena, including possibly both natural processes and industrial loadings.

WATER CHEMISTRY DATA FOR THE ST. CLAIR RIVER AT CORUNNA, ONTARIO, AS DETERMINED BY CONTINUOUS WATER QUALITY MONITOR

1. INTRODUCTION

Determining water quality by grab sampling techniques has inherent problems associated with defining a significance of the sample. Normally, numerous samples are collected to evolve a statistical representation. However, a valid statistical value for bodies of water which are generally both spatially and temporally variant may not be obtained from a few random samples. Nor are estimates of extremities likely to be available from such data.

Automatic water quality monitoring equipment which could measure and record various water quality parameters at fixed intervals was immediately recognized as being valuable in defining statistical and temporal relationships. Equipment of this type was used on the Potomac and Ohio rivers in the early 1960's (McDermott, 1969; Klein, Dunsmore and Horton, 1968) for water management purposes. However, to date, technology has been unable to provide the instrumentation required to measure many important water quality parameters. Nevertheless, automatic instruments can be used to record some of the simpler to measure parameters. It has thus been possible to record means, standard deviations, distributions or periodicities with respect to time, as well as extremities which would likely be missed during routine sampling. Even this information has been difficult to obtain, as is evidenced by the volume of equipment sales compared to the minute quantities of published data available. There is no doubt that the data are mandatory at least on a limited scale to design surveys and develop both statistical and mathematical water management models.

Submersible water quality recording meters have been operated in lakes Erie and Ontario (Palmer and Izatt, 1970; Palmer, 1972). These units have the advantages of portability and easy calibration and electronic repair in the laboratory. On the other hand, miniaturization results in a loss of reliability in the submersible meter. Calibrations cannot be adjusted in situ. Also, malfunction of any probe is not readily detected during the period of operation. A shore-based meter overcomes these difficulties, and in addition allows cleaning of the probes during operation and rapid calibration checking without significant down-time.

The St. Clair River was selected as a location for trial of a shore-based meter as it is influenced by the heavy density of chemical industries in the Sarnia area. To this end, a continuous water quality monitor with digital recording system (robot monitor) was operated from April 1971 to July 1972 at Corunna, downstream from the industrial area (Figure 1).

2. DESCRIPTION OF EQUIPMENT

The automatic water quality measuring system used in this study was manufactured by Schneider Instruments Co., who used the term "robot monitor" to describe the system (Model No. RM 25). The monitor measures chloride, temperature, dissolved oxyen (DO), pH and conductivity, and records these parameters periodically at a set sampling frequency on computer compatible punched paper tape. Although the sampling interval was generally 1 hour, the system could be set for several intervals between 10 minutes and 2 hours.

Temperature is measured by a silicon resistance element, which is incorporated in a DC bridge. The conductivity of the water between two platinum electrodes is measured with a 60 cycle AC bridge. Automatic temperature compensation produces a result corrected to 25°C. pH is measured with a glass electrode and silver-silver chloride reference electrode using a KCl solution as a salt bridge, and chloride with a similar system containing a chloride-sensitive glass electrode. DO is measured with a polarographic electrode, which consists of a gold cathode and four silver anodes immersed in a KCl solution, separated from the solution being analyzed by a membrane permeable to oxygen. Molecular oxygen passes from the water through the membrane, and is reduced at the cathode by electrons from the external voltage supply. These electrons are returned to the circuit by silver dissolution at the anodes, resulting in a current which is directly proportional to the number of oxygen molecules passing through the membrane.

During the period of this study, the robot monitor was housed at the Corunna sewage treatment plant. The intake was initially located at a depth of 1.5 m, approximately 10 m from shore, but as this was blocked by extensive weed growth in July 1971 and not very representative of the river, it was relocated about 30 m from shore at a depth of 3.5 m. A detailed description of the installation, along with problems encountered during the operation, is given in Appendix I.

3. STATISTICAL ANALYSIS OF DATA

Stochastic methods like time series and Markov chain analysis are valuable in processing large amounts of data produced by the robot monitor. Time series analysis detects periodic variations related to solar heating and photosynthesis, lake seiches and periodic waste loadings. A frequency and time distribution analysis defines the frequency of occurrence of extreme data values which may be detrimental to water use or aquatic life. Markov chain analysis can show the persistence of extreme values of parameters, i.e. will an extreme value remain present for a long period of time?

Prior to the application of these methods, data preparation is necessary. As the records contain errors derived from improper tape punching or false signals from the instrumentation or sensors, data editing is performed. For meaningful time series analysis, long term trends must be removed by trend correcting, in order that the mean value for each parameter may be applicable throughout the whole time period.

3.1 Tape Translation

The punched paper tape produced by the digitizer is read and the data transcribed onto magnetic tape. Dissolved oxygen figures are converted to percent saturation, using saturation data given by Standard Methods (APHA, 1965). Conductivity figures are standardized to 18°C using the Great Lakes composite conductivity-temperature data of Rodgers (1962). The data thus obtained are termed "raw data".

3.2 Data Editing

From the raw data, the mean and standard deviation of each chemical parameter (i.e. conductivity, pH, etc.) are determined. All observations greater or less than two standard deviations from the mean are rejected and replaced by linear interpolation between adjacent valid readings. This criterion allowed the rejection of nearly all the spurious data.

3.3 Trend Correcting

The entire edited record is divided into a number of equal sub-sets. These were approximate integer multiples of a day, in order to preserve diurnal variations (Palmer and Izatt, 1970). A polynomial is fitted to the means of these subsets against time, with the maximum order being two less than the number of subsets, or 5, whichever is less. A lower order is used if no improvement of fit occurs. A random scatter of the division means is poorly represented by polynomial fitting; consequently, the polynomials obtained are subjected to an F-test for significance of fit. If the polynomial is significant at 5 percent, a predicted value for the chemical parameter is obtained for each data point and subtracted from the reading. The difference is added to the grand mean of all the intervals to obtain a value which illustrates variations whose periods are less than the period of the subset.

3.4 Frequency and Time Distribution Analysis

The range of values obtained in each set of raw, edited and trend corrected data is divided into ten equal segments, and the frequency of observations falling within each segment determined. The frequency area moments are calculated. The distribution indicates the probability of obtaining high or low results, while the moments provide a measure of the mean, variance, skewness and kurtosis of the distribution (Snedecor and Cochran, 1967).

As many natural processes tend to produce a distribution which is normally or log-normally distributed with time, a measure of the probability of finding such distributions is desired. Four methods of calculating such a probability were tested. These include (1) the Kolmogorov-Smirnov test and (2) the Von Mises test applied to the continuous distribution of data points; as well as (3) the Von Mises discrete test and (4) the Chi-squared test applied to the frequency table already obtained. Data obtained during this comparison are presented in Appendix 3. As the result obtained for a theoretical normal distribution by the Von Mises continuous test differed from that expected by the authors, and the discrete tests gave variable results depending upon the manner in which frequencies were expressed, the Kolmogorov-Smirnov test was selected for use in time distribution analysis.

Both frequency and time distribution analysis are useful in assessing the magnitude of change obtained during data editing and trend correction.

3.5 Time Series Analysis

A time series analysis was conducted on the raw, edited and trend corrected data for each parameter to obtain the autocorrelation coefficients and the smoothed spectral density estimates. Analysis at each processing stage allows a comparison of raw, edited and trend corrected data to be made, and thus provides an estimate of the effects of each step on the data.

The time variation of the autocorrelation coefficient indicates the frequency of changes occurring in water chemistry parameters, and aids in the interpretation of such changes. Spectral analysis improves the resolution of complex periodicities and is useful when the auto-correlation function shows several interfering frequencies.

For meaningful time series analysis, data must be "pre-whitened" or smoothed to remove irregular short-term fluctuations. Each smoothed data point is obtained by forming a running mean of several consecutive observations. Although an equally weighted running mean is simplest to calculate, in order to obtain a smoothly varying frequency response, the observations are weighted proportionally to the binomial coefficients, resulting in a binomially weighted running mean (Panofsky and Brier, 1968, p. 147-151). In general, hourly data were pre-whitened over 5 time intervals allowing the observation of spectral peaks at all periods of interest.

Autocorrelation coefficients r_L of the time series formed by each of the parameters are computed for lag time L up to 100 hours with the equation:

$$r_L = \sum_{i=1}^{N} \frac{(X_i - \bar{X})\ (X_{i+L} - \bar{X})}{S_x{}^2 N} \qquad \ldots (1)$$

where X_i is the ith reading of any parameter

$\bar{X}$ is the mean of N readings of X_i

$S_x{}^2$ is the variance of N readings of X_i

Spectral density estimates are obtained by Fourier cosine transformation of the autocorrelation function (Panofsky and Brier, 1968, p. 143). The spectral estimates B_i are obtained from:

$$B_i = \frac{k}{m} \left(r_o + r_m \cos (i\pi) + 2\sum_{L=1}^{m-1} r_L \cos(Li\pi/m) \right) \ldots (2)$$

where r_o, r_L and r_m are autocorrelation coefficients for lags 0, L and m (m = maximum lag time).

L and i are lag times, each varying individually from 0 to m.

k is a constant, k = 1/2 for i = 0 and i = m

k = 1 for other values of i.

As equation (2) produces rapidly fluctuating spectral density estimates, smoothed spectral estimates are computed by a technique called Hanning (Blackman, 1958), according to:

$$S_i = 0.25\ B_{i-1} + 0.50\ B_i + 0.25\ B_{i+1} \qquad \ldots (3)$$

where S_i = smoothed spectral estimate.

These estimates are plotted against frequency $\frac{i}{2m\Delta t}$

where Δt is the measuring interval
(in most cases, 1 hr) and i and m are as in (2).

In order to determine the statistical significance of peaks in the spectral estimate curve, an average line was drawn through the points, and the ratios of maxima and minima to this average were calculated. The limiting levels of significance for these maxima and minima may then be estimated from the plot given by Jenkins (1968, p.82), where the number of degrees of freedom ν is estimated from $\nu = (2N - m/2)/m$ where N and m are the total number of readings and maximum lag time, as stated previously.

3.6 Markov Chain Analysis

The range between the minimum and maximum readings for each chemical parameter is divided into ten states as in Section 3.4, and these are numbered consecutively from one to ten, one being the state containing the lowest readings. A 10 x 10 one hour transition probability matrix $T_{mn}(1)$ (where m = initial state and n = state after 1 hour) is determined for the period of record. This matrix defines the probability of occurrence of each of the ten possible states after one hour, given the present state. From this matrix, it is possible to determine the final state probability vector after any number (k) of hours, assuming that the initial state probability vector for k hours is equal to the final state vector for (k-1) hours. By comparing the probability after k hours with the final state probability vector, one may estimate the time required for the system to return to an average state following the recording of an extreme value.

The final state probability vector is the probability of each state occurring over the period of the record, as determined by frequency analysis (Sec. 3.4). Probability vectors for k hours may be determined by the transition matrix using the relation:

$$q_n(k) = \sum_{m=1}^{10} q_m(k-1)T_{mn}(1) \qquad \ldots(4)$$

where

$q_n(k)$ = final state probability vector after k hours

$q_m(k-1)$ = probability vector after k-1 hours or initial state probability vector for k = 1

$T_{mn}(1)$ = one hour probability transition matrix

The initial state is usually defined at an extreme reading (e.g. state 1, probability = 1.0), and probability vectors for 1-40 hours are determined for each chemical parameter. A rapid progression to the final state frequency vector inidcates that the extreme values are rare and unlikely to persist for long periods of time, while a slow progression indicates the values vary greatly and that extremes are not unlikely.

4. RESULTS AND DISCUSSION

The robot monitor was operated continuously from April 1971 to July 1972. Readings were made at hourly intervals, except in April 1971 when a two hour interval was used.

In many cases, equipment malfunctioning resulted in invalid data. These included the action of weeds and ice, as well as calibration difficulties and digitizer failures. In August-September 1971, the paper tape was punched off center, and could not be translated. In May-July 1972, increasing electronic difficulties in the digitizer resulted in bad data being obtained. These problems are discussed in detail in Appendix 1. A complete record of all data tapes obtained is given in Appendix 2.

The only periods in which reasonable results were obtained were April-June 1971 and November 1971-February 1972, except for a period in late November, when several sensors were out of calibration. Other time periods will not be discussed further.

4.1 Data Preparation

Complete listings of raw, edited and trend corrected data are routinely printed so that the effect of each processing step on the data record may be ascertained. Table 1 illustrates a sample listing of data from April 1971.

Tables 2-4 and Figures 2-4 illustrate the frequency and time distribution analyses for raw, edited and trend corrected chloride data from April 1971. The removal of mispunched data on the paper tape by editing is illustrated. As such mispunching produces either very high or very low results, editing often produces a considerable change. In this case, the range of chloride values is seen to decrease from 20-325 mg/l for raw data, to 20-58 mg/l for edited data. This decrease is also accompanied by a decrease in the moments of the distribution, as shown in Tables 2 and 3.

The results of polynomial fitting to the edited data for purposes of trend correction are given in Table 5. Typical plots of polynomials obtained during trend correction have been published previously (Palmer and Izatt, 1970). In several cases, the number of subsets was varied in order to determine whether or not any bias occurred in the data from use of a given number of subsets. This factor is illustrated in Table 5 for December 1971. The results are similar for either 10 subsets (approximately 3 days each) or 14 subsets (approximately 2 days each), although increasing the number of subsets tends to increase the level of significance.

The effect of trend correcting on the frequency and time distribution analysis of chloride data from April 1971 is illustrated in Tables 3-4 and Figures 3-4. Trend correcting results in a further decrease in the range of data to 22-42 mg/l. While the practice of trend correction is necessary for time series analysis, it should be approached with caution. This decrease in range, together with removal of skewness and increased probability of normal distribution, actually results in the submergence of extreme values of data, which are important ecologically. An example of this is the chloride data discussed in Section 4.4.

Means and standard deviations of all data thus obtained are presented in Table 6. Standard deviations obtained by robot monitor for chloride, DO and conductivity are similar to or slightly lower than standard deviations of six grab samples obtained between May and November, 1971. As a fraction of the mean, standard deviations of the warm and cold weather data are similar, except for conductivity, where they are higher during the cold weather period.

4.2 Frequency and Time Distribution Analysis

Results of the Kolmogorov-Smirnov test for normal, log-normal or exponential distribution are presented in Table 7. Distribution is an important characteristic since the application of parametric statistical tests to the data assumes normal distribution. If a large number of processes, natural or artificial, are combining to produce effects, a normal or log-normal distribution is expected. For example, conductivity is influenced by all ionic species present in the water, which in turn may be affected by a considerable number of factors such as municipal and industrial discharges, rainfall, biological activity, wave and current action. Non-normally distributed properties, on the other hand, frequently result when one or two factors are primarily responsible for observed variations.

In general, cold weather data appear to be more normally distributed than warm weather data, particularly in the case of conductivity and dissolved oxygen, which are strongly affected by photosynthesis in the growing season. Distributions for chloride and temperature were unpredictable and the pH distribution was generally not normal.

4.3 Time Series Analysis

The results of time series analysis, presented in Table 8, reflect the difference between the warm and cold weather records, with warm weather data indicating diurnal periodicities attributable to solar heating and photosynthesis and cold weather data generally being less periodic in nature. Typical autocorrelation functions from June 1971 are presented in Figures 5-8 and the variance density spectra derived from these in Figures 9-12.

4.3.1 Warm Weather Data

The diurnal effect of solar heating is obvious in all warm weather data. Maximum temperatures are generally recorded at 1400 - 2000 hr, and minima at 0200 - 0800 hr, with the average daily amplitude of variation between 1 and 2°C.

Photosynthetic activity should result in maximum DO and maximum pH late in the day (due to the conversion of dissolved CO_2 to O_2). Plant growth can occur to depths at which one percent of the incident sunlight has not been absorbed; this is equivalent to twice the secchi disk reading, or about 3 to 6 m depth for the St. Clair River. Considering this fact, photosynthetic activity should be easily seen for this time period, as the intake was at a depth of only 1.5 m.

This is apparent from the spectral peaks for pH in May and DO in June (Fig. 11). Less pronounced diurnal variations in DO are also evident in April, with a typical amplitude of change being 5 percent in April and up to 25 percent in June. This magnitude of change in percent saturation is too large to be explained by temperature changes. These observations indicate the effect of rapid weed growth on the DO.

Variations in the conductivity record are less regular. In a portion of the April record, maximum conductivity and minimum DO tended to complement each other, but at other times the time variation of conductivity was irregular. These conductivity variations could be due to regular waste loadings, causing increased conductivity and oxygen demand, but this was not continuously observed. In June, a significant semi-diurnal periodicity of conductivity is obvious on examining the spectral plot (Figure 12). However, this periodicity is not present in the autocorrelation function (Figure 8). The reason for this periodicity is not known.

Calcium and bicarbonate ions are the major contributors to the conductivity as evidenced by analyses obtained by the Great Lakes Surveys (Table 9). This is generally true for most fresh waters (Hutchinson, 1957, Chapter 9). Data for this area of the St. Clair River from the Great Lakes Surveys indicate that industrial loadings result in increased chloride concentrations near the shore, while midstream data are typical of Lake Huron water. The effect of this chloride loading on the conductivity is seen in Table 9. The importance of sulfate ion in determining conductivity is not known in this case, as analytical data are not available for this ion.

As the major portion of chloride is due to industrial loadings, any periodicity in this parameter will reflect periodicities in loading. However, no periodicities are observed for chloride at the 80 percent confidence level (e.g. Figure 9). If the semi-diurnal conductivity variations were due to chloride loadings, a strong semi-diurnal variation should have been observed. As this was not the case, the source of the conductivity variation is not known.

4.3.2 Cold Weather Data

The absence of biological growth during this period results in an almost complete absence of diurnal variations. Apparent exceptions are DO in December 1971 and conductivity in January 1972. However, the complete spectral plot of DO in December indicates a large number of peaks below the 80 percent confidence level as well as the indicated 24 and 10 hour peaks. Interpretations of complex spectra of this nature must be approached with caution as they may simply represent white noise generated by the analyses (Jenkins and Watts, 1968, p. 157 and 230-9). The observed periodicities for pH and conductivity in December can be similarly explained.

In the case of January conductivity, a tendency towards maximum late in the day and minimum early in the day suggests conductivity loading during daylight hours. As stated earlier, conductivity is mainly due to bicarbonate but chloride from industrial wastes is also important. However, no similar trend exists in the chloride data. Sulfate may be important here but analytical data are not available.

Chloride data tend to be random. For January and February 1972, most of the variance is accounted for by periods which are close to zero, where the refined spectral estimate values are 1 to 2 orders of magnitude greater than the height of the observed peaks. The absolute value of the spectral estimate, as well as the size of the peak is certainly important in determining its significance, and these peaks cannot be regarded as significant.

The 8-hour periodicity observed for DO in November 1971 is evident upon examining both the data record and the autocorrelation function. The amplitude of variation is frequently 10 percent or more. Some evidence of a 4-hour period is also observed, though not at a significance level of 80 percent. The non-normal time distribution of DO is also significant. (DO tended to be normally distributed in December and January). This could conceivably suggest BOD or COD loading, but other evidence is required. Evidence of this nature would be substantiated if data were available from several monitors at different locations. This would allow a more detailed picture of the waste loadings to the river.

In summary, the periodicities observed during cold weather are completely unlike what is seen during the growing season, as is to be expected. They tend to vary from month to month, and are accompanied by random short-term changes. In many cases, the variance density spectrum is concentrated near zero frequencies, suggesting that changes are random, and presumably due to a combination of factors, both natural and artificial.

4.4 Examples of Extreme Data Observed

One important purpose of automatic water quality monitoring equipment is to detect extreme results not ordinarily detected by routine sampling. This was realized during the spring of 1971, when evidence suggesting runoff of salt from city streets was obtained. A short period of low dissolved oxygen values was also seen. In the following discussion, quoted values refer to edited data as the process of trend correction tends to obliterate the extremes.

The chloride level was highest at the beginning of the April record (50 mg/l) and decreased thereafter with the June mean being 23 mg/l. A possible cause of the April maximum is the discharge of salt used on Sarnia streets to the river, following a rainfall. Overflow from the combined storm and sanitary sewers may also be important. Meteorological observations (Canada Department of Transport, 1971) indicate that 0.44 inches of rain fell at Sarnia on April 12-13, two days before the beginning of the record. An 0.29 inch rainfall on the 17th, however, was not followed by a chloride maximum. An 0.16 inch rainfall on the 28th following a snowfall on the 24th, was followed by a short duration pulse of increased chloride on the 30th, suggesting runoff of chloride from city streets with a 2-day lag time. Increased chloride content in Toronto Harbour following runoff after a snowfall has been demonstrated (Brydges, 1969) and increased chloride has been evident in Hamilton Harbour following spring runoff for several years. It is concluded that there is considerable evidence for runoff of chloride affecting results at the robot monitor, but data are not entirely consistent.

In May 1971, an unusual decrease in dissolved oxygen from 100 to 55 percent saturation was observed on May 23-24. This was followed by an increase to 75-80 percent saturation on May 25-26, and 100 percent at the end of the record (May 27), when calibration indicated no significant error. A comparison of edited DO frequency distributions for April-June 1971, given in Figure 13, illustrates the effect of this change on the frequency analysis.

This is another good example of extreme data which would likely be missed during routine surveys. However, no correlation with available meteorological data existed, and no explanation for this drop in DO is available.

4.5 Markov Chain Analysis

In order to obtain meaningful results from this technique, data selected should have a sufficiently large range from minimum to maximum that the width of each state will be greater than the usual precision of measurement for each parameter (Appendix 1, Section A1.4). To this end, edited data for chloride, dissolved oxygen and conductivity from June and December 1971 were analyzed. The results are shown in Figures 14-16.

The probabilities of occurrence of various states for these parameters were determined for each hour up to 40 hours, given an initial reading within state 1. These probabilities provide an indication of the hour-by-hour variation of these parameters. For example, if the dissolved oxygen was 80 percent saturated at a given time (State 1, Figure 15) what is the probability of it being 120 percent saturated four hours later? The observed probability is close to zero, indicating that changes of this magnitude are not observed in this length of time.

In the case of June 1971, figures 14-16 show that, following an initial reading in state 1, the state probability vector approaches the final state probability vector most rapidly for chloride, a good approximation being achieved after 12 hours. (Composite sampling would provide the same information). However, with conductivity, an extremely slow progression is noted, with the probability of being in state 1 still above 0.5 at 4 hours. Even after 24 hours, a good approximation has not yet been achieved. This indicates the persistence of extreme values of conductivity over periods of even greater than one day. If an automatic warning system were established to indicate such an extreme, adequate time is available to obtain a sample for further analysis to elucidate the reason for such a reading. With dissolved oxygen, a good approximation is achieved after 12 hours, as can be expected from the diurnal variation of this parameter. The 4 and 8 hour curves indicate some tendency for abnormally low values to persist for several hours.

In contrast to June 1971, an extremely rapid progression to the final state was noted for the December 1971 conductivity data. (The state vector for 4 hours was a very close approximation to the final state vector.) This presumably contrasts the effect of photosynthesis on the June data. On the other hand, the rates of progression for chloride and dissolved oxygen were similar in the two records, although a faster rate of progression might have been expected for the December dissolved oxygen, considering the lack of photosynthesis in this time period.

ACKNOWLEDGEMENTS

Assistance was provided by J.B. Izatt of the Systems and EDP Branch, who did much of the computer programming associated with this study. Periodic maintenance and calibration of the equipment was performed by J. Gabay and D. Kennedy of the Water Quality Branch, and S. Bechard of the Corunna STP. Finally, Dr. T.G. Brydges of the Chemistry Branch read the manuscript and made several valuable suggestions for its improvement.

REFERENCES

APHA, 1965. Standard methods for the examination of water and wastewater. American Public Health Assoc., New York. 769 pp.

BLACKMAN, R.B. and J.W. TUKEY, 1958. The measurement of power spectra. Dover Publications, New York, 190 pp.

BRYDGES, T., 1969. Investigation of daily variations in chemical, bacteriological and biological parameters at two Lake Ontario locations near Toronto. Part 1 - Chemistry. Proc. 12th Conf. Great Lakes Res., Int. Assoc. Great Lakes Res., 750-759.

CANADA DEPARTMENT OF TRANSPORT, 1971. Meteorological Branch, daily climatological summary, listing no. 6127514 534. 2 pp.

HUTCHINSON, G.E., 1957. A treatise on limnology. John Wiley and Son, New York. Vol. 1.

JENKINS, G.M. and D.G. WATTS, 1968. Spectral analysis and its applications. Holden-Day, San Francisco. 525 pp.

KLEIN, W.L., D.A. DUNSMORE, and R.K. HORTON, 1968. An integrated monitoring system for water quality management in the Ohio Valley. Env. Sci. Technol. 2:764-771.

McDERMOTT, J.H., 1969. Water quality monitoring. Presented at the Environmental Engineering Conference, University of Florida, Gainesville, Florida.

PALMER, M.D., 1972. Some chemical and physical relationships on Lake Ontario. Water Research 6:843-858.

PALMER, M.D. and J.B. IZATT, 1970. Determination of some chemical and physical relationships from recording meters in lakes. Water Research 4:773-786.

PANOFSKY, H.A. and G.W. BRIER, 1968. Some applications of statistics to meteorology. The Pennsylvania State University, University Park, Pa., 224 pp.

RODGERS, G.K., 1962. Temperature adjustment of conductivity measurements in waters of the Great Lakes. Great Lakes Institute, University of Toronto. 7 pp.

SNEDECOR, G.W. and W.G. COCHRAN, 1967. Statistical methods, 6th ed. Iowa State University Press, Ames, Ia. 593 pp.

```
                                   CODE  81901      EDITED
   RECORD        2                                                                      PAGE     2

  DATE          TIME        CL     TEMP(CENT)    D.O.(%)   TURB(JCU)     P7          PH      COND(18 C)      CYCLE

 180471.        14.00      34.00       6.89      112.79   777777.00   777777.00      7.97      252.96
 180471.        16.00      33.00       6.56      113.44   777777.00   777777.00      8.30      254.35
 180471.        18.00      31.50       6.33      110.88   777777.00   777777.00      8.34      253.00
 180471.        20.00      30.00       6.17      110.24   777777.00   777777.00      8.21      252.36
 180471.        22.00      32.00       6.11      109.44   777777.00   777777.00      8.07      254.23
 190471.         0.0       33.50       6.00      109.60   777777.00   777777.00      8.05      255.30
 190471.         2.00      28.50       5.94      110.08   777777.00   777777.00      8.05      246.43
 190471.         4.00      30.50       5.83      110.08   777777.00   777777.00      8.08      249.29
 190471.         6.00      32.00       5.94      109.60   777777.00   777777.00      8.09      251.81
 190471.         8.00      28.50       6.00      110.88   777777.00   777777.00      8.09      247.24
 190471.        10.00      28.00       6.17      111.68   777777.00   777777.00      8.17      246.99
 190471.        12.00      29.50       6.72      114.59   777777.00   777777.00      8.11      248.73
 190471.        14.00      39.50       7.06      114.92   777777.00   777777.00      8.04      258.07
 190471.        16.00      34.00       7.00      115.25   777777.00   777777.00      8.11      255.48
 190471.        18.00      31.50       6.67      115.57   777777.00   777777.00      8.28      252.39
 190471.        20.00      31.25       6.28      113.95   777777.00   777777.00      8.17      248.61
 190471.        22.00      31.00       6.11      112.32   777777.00   777777.00      8.07      247.97
 200471.         0.0       31.00       6.17      111.84   777777.00   777777.00      8.08      248.78
 200471.         2.00      31.00       5.89      111.68   777777.00   777777.00      8.06      247.41
 200471.         4.00      31.00       5.83      112.00   777777.00   777777.00      8.05      245.71
 200471.         6.00      31.50       5.89      111.36   777777.00   777777.00      8.03      246.52
 200471.         8.00      30.00       6.11      111.68   777777.00   777777.00      8.06      246.18
 200471.        10.00      27.00       6.33      112.32   777777.00   777777.00      7.96      240.49
 200471.        12.00      28.00       6.72      115.41   777777.00   777777.00      8.06      244.26
 200471.        14.00      27.50       7.06      115.57   777777.00   777777.00      8.08      244.68
 200471.        16.00      23.50       6.83      116.23   777777.00   777777.00      8.06      237.84
 200471.        18.00      28.50       6.67      115.57   777777.00   777777.00      8.23      249.71
 200471.        20.00      33.00       6.50      112.48   777777.00   777777.00      8.21      258.02
 200471.        22.00      27.50       6.39      112.48   777777.00   777777.00      8.12      245.77
 210471.         0.0       30.00       6.50      111.68   777777.00   777777.00      8.12      249.96
 210471.         2.00      31.50       6.39      111.36   777777.00   777777.00      8.16      252.92
 210471.         4.00      30.00       6.28      110.72   777777.00   777777.00      8.19      249.51
 210471.         6.00      28.50       6.33      109.60   777777.00   777777.00      8.20      246.74
 210471.         8.00      28.00       6.44      109.12   777777.00   777777.00      8.05      247.48
 210471.        10.00      26.00       6.50      110.40   777777.00   777777.00      8.01      245.48
 210471.        12.00      23.50       6.33      110.56   777777.00   777777.00      8.14      244.96
```

TABLE 1   Sample Robot Monitor Data

```
                                   CHLORIDE

                                FREQUENCY TABLE

                        CLASSES                          FREQUENCY(%)
                      20.00 -      50.50                 96.93
                      50.50 -      81.00                  1.75
                      81.00 -     111.50                  0.0
                     111.50 -     142.00                  0.0
                     142.00 -     172.50                  0.0
                     172.50 -     203.00                  0.0
                     203.00 -     233.50                  0.0
                     233.50 -     264.00                  0.44
                     264.00 -     294.50                  0.0
                     294.50 -     325.00                  0.88

                   TOTAL                                  100.



      MOMENTS
         1          39.129
         2          29.363
         3          59.863
         4          86.322



      AVERAGES FOR THE THIRDS

      FIRST THIRD =      38.470

      SECOND THIRD =     32.477

      THIRD THIRD =      30.954


    THE HYPOTHESIS THAT THE SAMPLE IS FROM A(N)  NORMAL    DISTRIBUTION
     WITH MEAN       33.9671 AND VARIANCE     945.3906
     CAN BE REJECTED WITH PROBABILITY 0.0   OF BEING INCORRECT.  THE STATISTIC Z
     IS  0.5346E 01 FOR THIS SAMPLE.


    THE HYPOTHESIS THST THE SAMPLE IS FROM A(N) EXPONENTIAL DISTRIBUTION
     WITH MEAN       33.9671 AND VARIANCE     945.3906
     CAN BE REJECTED WITH PROBABILITY 0.0   OF BEING INCORRECT.  THE STATISTIC Z
     IS  0.6769E 01 FOR THIS SAMPLE.


    THE HYPOTHESIS THST THE SAMPLE IS FROM A(N)  LOG NORMAL DISTRIBUTION
     WITH MEAN        1.4904 AND VARIANCE        0.0184
     CAN BE REJECTED WITH PROBABILITY 0.0   OF BEING INCORRECT.  THE STATISTIC Z
     IS  0.3321E 01 FOR THIS SAMPLE.
```

TABLE 2   Raw Data, April 1971

```
                                                    CHLORIDE

                                                FREQUENCY TABLE

                                    CLASSES                                FREQUENCY(%)
                                   20.00 -     23.80                         7.46
                                   23.80 -     27.60                        24.12
                                   27.60 -     31.40                        37.72
                                   31.40 -     35.20                        17.11
                                   35.20 -     39.00                         5.26
                                   39.00 -     42.80                         3.95
                                   42.80 -     46.60                         0.88
                                   46.60 -     50.40                         1.32
                                   50.40 -     54.20                         1.32
                                   54.20 -     58.00                         0.88

                                 TOTAL                                     100.



      MOMENTS
         1         30.433
         2          6.114
         3          7.253
         4          9.921



      AVERAGES FOR THE THIRDS

      FIRST THIRD =      34.648

      SECOND THIRD =     28.727

      THIRD THIRD =      28.197

   THE HYPOTHESIS THAT THE SAMPLE IS FROM A(N)  NORMAL   DISTRIBUTION
    WITH MEAN       30.5241 AND VARIANCE       34.2983
    CAN BE REJECTED WITH PROBABILITY 0.000 OF BEING INCORRECT.  THE STATISTIC Z
    IS  0.2604E 01 FOR THIS SAMPLE.


   THE HYPOTHESIS THST THE SAMPLE IS FROM A(N) EXPONENTIAL DISTRIBUTION
    WITH MEAN       30.5241 AND VARIANCE       34.2983
    CAN BE REJECTED WITH PROBABILITY 0.002 OF BEING INCORRECT.  THE STATISTIC Z
    IS  0.1854E 01 FOR THIS SAMPLE.


   THE HYPOTHESIS THST THE SAMPLE IS FROM A(N)  LOG NORMAL DISTRIBUTION
    WITH MEAN        1.4774 AND VARIANCE        0.0059
    CAN BE REJECTED WITH PROBABILITY 0.001 OF BEING INCORRECT.  THE STATISTIC Z
    IS  0.1974E 01 FOR THIS SAMPLE.
```

TABLE 3 Edited Data, April 1971

```
                                        CHLORIDE

                                     FREQUENCY TABLE

                         CLASSES                               FREQUENCY(%)
                        21.72 -      23.69                      1.32
                        23.69 -      25.67                      5.26
                        25.67 -      27.65                     11.84
                        27.65 -      29.63                     21.49
                        29.63 -      31.61                     22.81
                        31.61 -      33.58                     22.81
                        33.58 -      35.56                      7.89
                        35.56 -      37.54                      3.51
                        37.54 -      39.52                      1.32
                        39.52 -      41.50                      1.75

                     TOTAL                                     100.


     MOMENTS
        1          30.557
        2           3.357
        3           2.352
        4           4.574


     AVERAGES FOR THE THIRDS

     FIRST THIRD =       30.426

     SECOND THIRD =      30.820

     THIRD THIRD =       30.377

THE HYPOTHESIS THAT THE SAMPLE IS FROM A(N)  NORMAL   DISTRIBUTION
 WITH MEAN       30.5401 AND VARIANCE       11.3331
 CAN BE REJECTED WITH PROBABILITY 0.673 OF BEING INCORRECT.  THE STATISTIC Z
 IS  0.7228E 00 FOR THIS SAMPLE.


THE HYPOTHESIS THST THE SAMPLE IS FROM A(N) EXPONENTIAL DISTRIBUTION
 WITH MEAN       30.5401 AND VARIANCE       11.3331
 CAN BE REJECTED WITH PROBABILITY 0.000 OF BEING INCORRECT.  THE STATISTIC Z
 IS  0.2517E 01 FOR THIS SAMPLE.


THE HYPOTHESIS THST THE SAMPLE IS FROM A(N)  LOG NORMAL DISTRIBUTION
 WITH MEAN        1.4822 AND VARIANCE        0.0023
 CAN BE REJECTED WITH PROBABILITY 0.752 OF BEING INCORRECT.  THE STATISTIC Z
 IS  0.6753E 00 FOR THIS SAMPLE.
```

TABLE 4   Trend Corrected Data, April 1971

TABLE 5

POLYNOMIAL FITTING DURING TREND CORRECTION

Data Period	Number of Intervals	Level of Significance from F-Test (percent)				
		Cl^-	Temp.	DO	pH	Cond.
April 1971	9	1.0	NS	1.0	NS	NS
May 1971	7	-	5.0	*	NS	NS
June 1971	13	NS	0.1	5.0	-	5.0
early Nov. 1971	11	-	0.1	NS	-	2.5
December 1971	14	NS	0.5	0.1	NS	-
	10	NS	2.5	0.5	NS	NS
January 1972	6	*	*	NS	NS	NS
February 1972	13	1.0	NS	-	NS	2.5

Note: * Although the polynomial was not significant at 5%, trend correction was accepted because the portion of the variance attributed to deviation about regression was less than 2% of the total variance.

NS indicates that the polynomial fit was not significant at the 5% level.

The chloride probe was out of calibration in May and early November 1971. The pH reference probe was malfunctioning in June 1971 and calibration in November was questionable. The dissolved oxygen probe was out of calibration in February 1972.

TABLE 6

MEANS AND STANDARD DEVIATIONS OF CORUNNA ROBOT MONITOR DATA

Data Period	Cl^- (mg/l)		Temp. (°C)		Dissolved Oxygen (% saturation)		pH		Conductivi (µmho/cm at 1	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.
April 1971	30.5	3.7	6.3	0.4	105.7	2.2	8.2	0.1	253	8
May 1971	-	-	10.7	0.7	83.1	10.1	8.1	0.3	317	9
June 1971	22.8	6.1	13.0	0.6	103.8	5.8	-	-	310	7
early November 1971	-	-	12.8	1.0	82.3	5.6	-	-	224	9
December 1971	32.0	6.0	6.2	1.1	84.5	4.9	8.1	0.3	235	13
January 1972	29.2	8.9	1.6	0.8	74.0	4.8	7.7	0.2	228	22
February 1972	35.0	8.7	1.5	0.7	-	-	7.8	0.3	241	14

Note: Trend removed data are used where significant, otherwise edited data.

Readings were obtained at hourly intervals, except in April 1971, when the interval was 2 hours.

TABLE 7

DISTRIBUTION PROBABILITIES BY KOLMOGOROV-SMIRNOV TEST

Data Period	Cl^-		Temp.		DO		pH		Conductivity	
	Dist.	Prob.	Dist.	Prob.	Dist.	Prob.	Dist.	Prob.	Dist.	Prob.
April 1971	log norm.	.752	log norm.	.055	normal	.229	log norm.	.306	exp.	.016
May 1971	-	-	normal	.000	log norm.	.172	normal	.0	log norm.	.184
June 1971	log norm.	.012	normal	.632	normal	.000	-	-	normal	.000
early Nov. 1971	-	-	log norm.	.002	normal	.094	-	-	normal	.453
Dec 1971	normal	.025	normal	.0	normal	.246	normal	.0	log norm.	.994
Jan 1972	normal	.790	log norm.	.527	normal	.836	normal	.002	log norm.	.830
Feb 1972	normal	.045	normal	.000	-	-	normal	.0	normal	.405

Note: Trend removed data has been used where significant, otherwise edited. The distribution with the highest probability is indicated.

TABLE 8

80% CONFIDENCE SPECTRAL MAXIMA

Data Period	No. of Data Points	No. of Lags	Cl^-	Temp.	DO	pH	Cond.
April 1971	228	30	None	*24 hr 12 hr	(24 hr)	(24, 12 hr)	(24, 12 hr)
May 1971	162	30	-	24 hr	None	24 hr	None
June 1971	625	100	None	*24 hr	*24 hr	-	12 hr
early Nov. 1971	525	30	-	None	8 hr	-	None
Dec 1971	689	100	None	None	24 hr 10 hr	10 hr	16 hr 6.3 hr
Jan 1972	144	30	5 hr	None	None	None	24 hr
Feb 1972	900	30	10 hr	None	-	None	None

Notes:

* indicates that the peak is significant at the 95% confidence level.

( ) indicates interesting periodicities which were visible but below the 80% level.

Data were prewhitened by binomial weighted running mean over 5 hourly observations.

TABLE 9

SOME ANALYTICAL DATA FOR ST. CLAIR RIVER DETERMINED BY GREAT LAKES SURVEYS

This data is for Station SR-31.0, 42-53-54N, 82-28-18W, 1 Km north of the robot monitor intake.

Date	Location	Conductivity (25°C) µmhos	Cl^- mg/l	Alkalinity mg $CaCO_3$/l
May 5/71	M	213	7	-
	S	274	26	-
May 31/71	M	209	5	81
	S	256	19	92
June 28/71	M	212	6	89
	S	252	17	88
Oct. 26/71	M	209	7	84
	S	276	24	88

Note: M = midstream (640 m from U.S. shore)

S = Canadian shore (1,130 m from U.S. shore)

Average equivalent concentrations:			
	Cl^-	(M)	0.2 meq/l
		(S)	0.6 meq/l
	Alk		1.7 meq/l

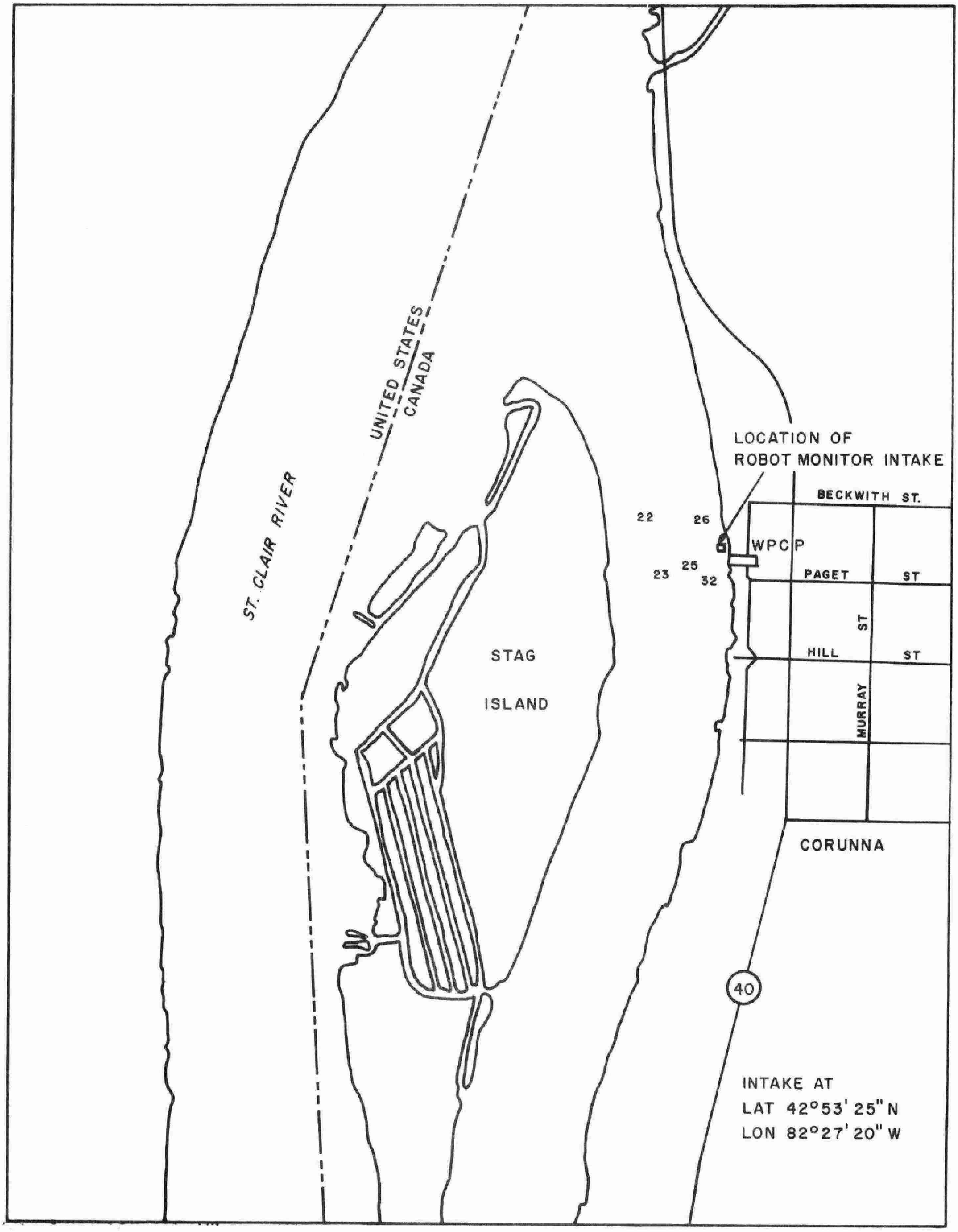

Fig. 1 Location of robot monitor

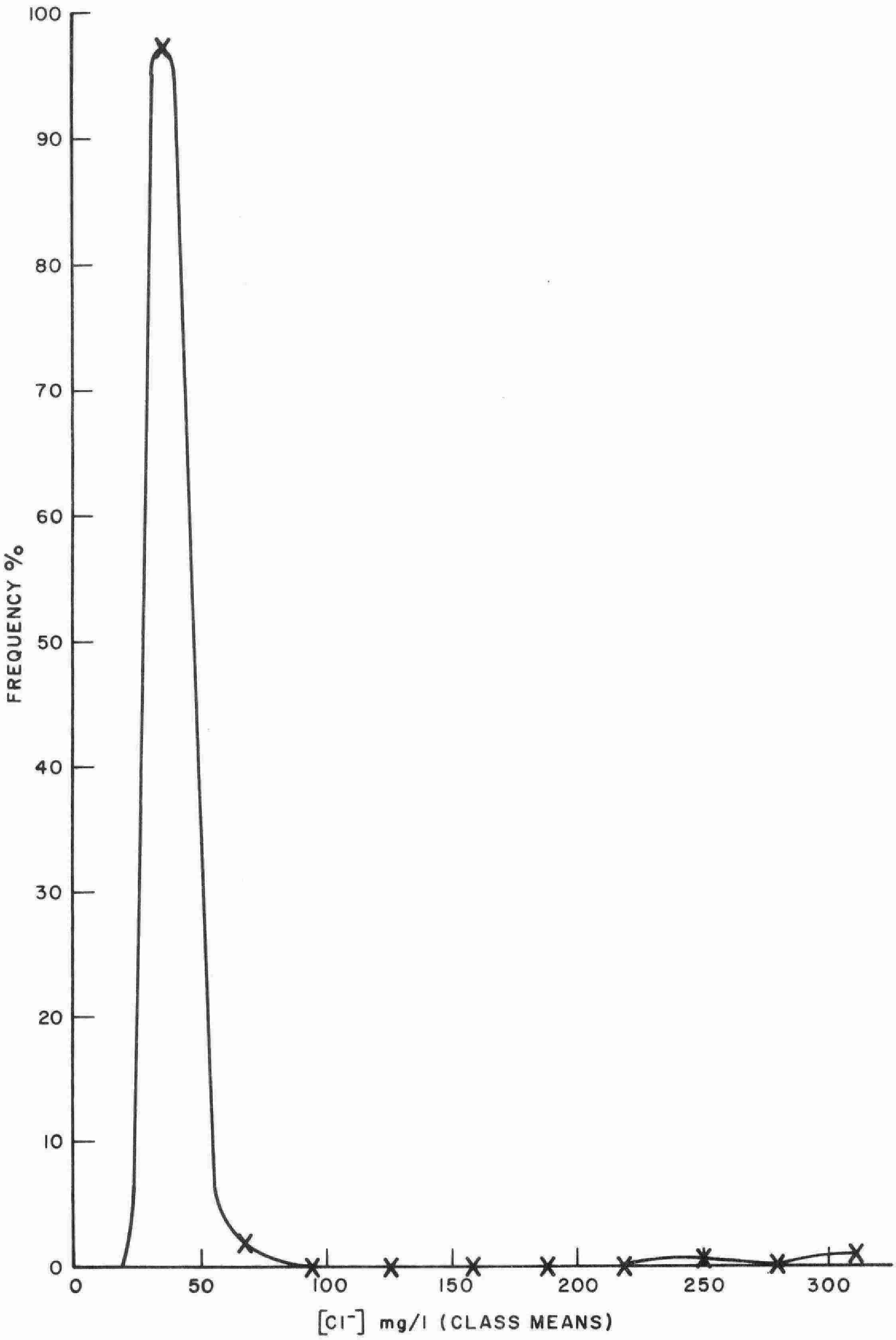

Fig. 2 Raw chloride, April 1971, frequency distribution

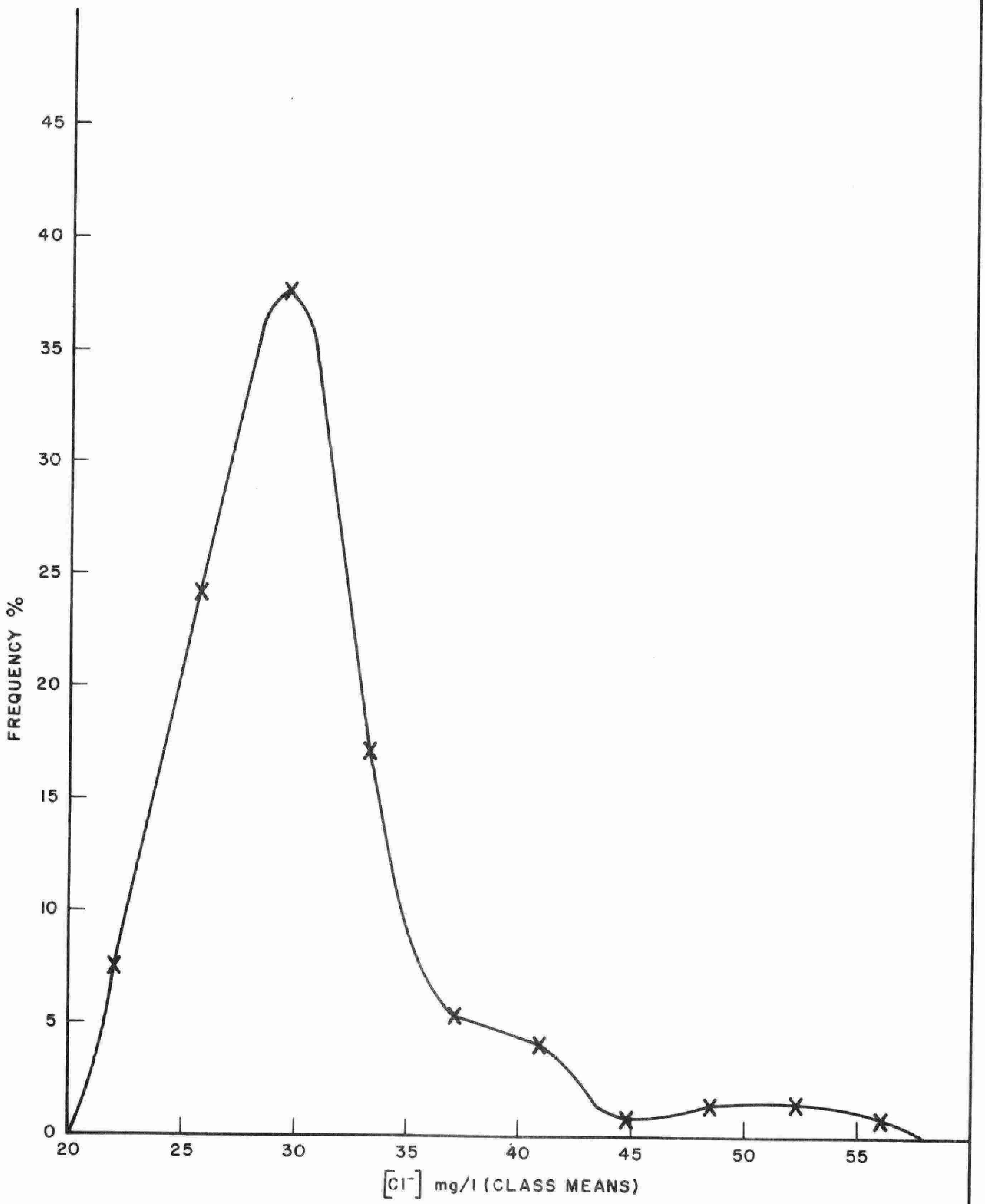

Fig. 3 Edited chloride, April 1971, frequency distribution

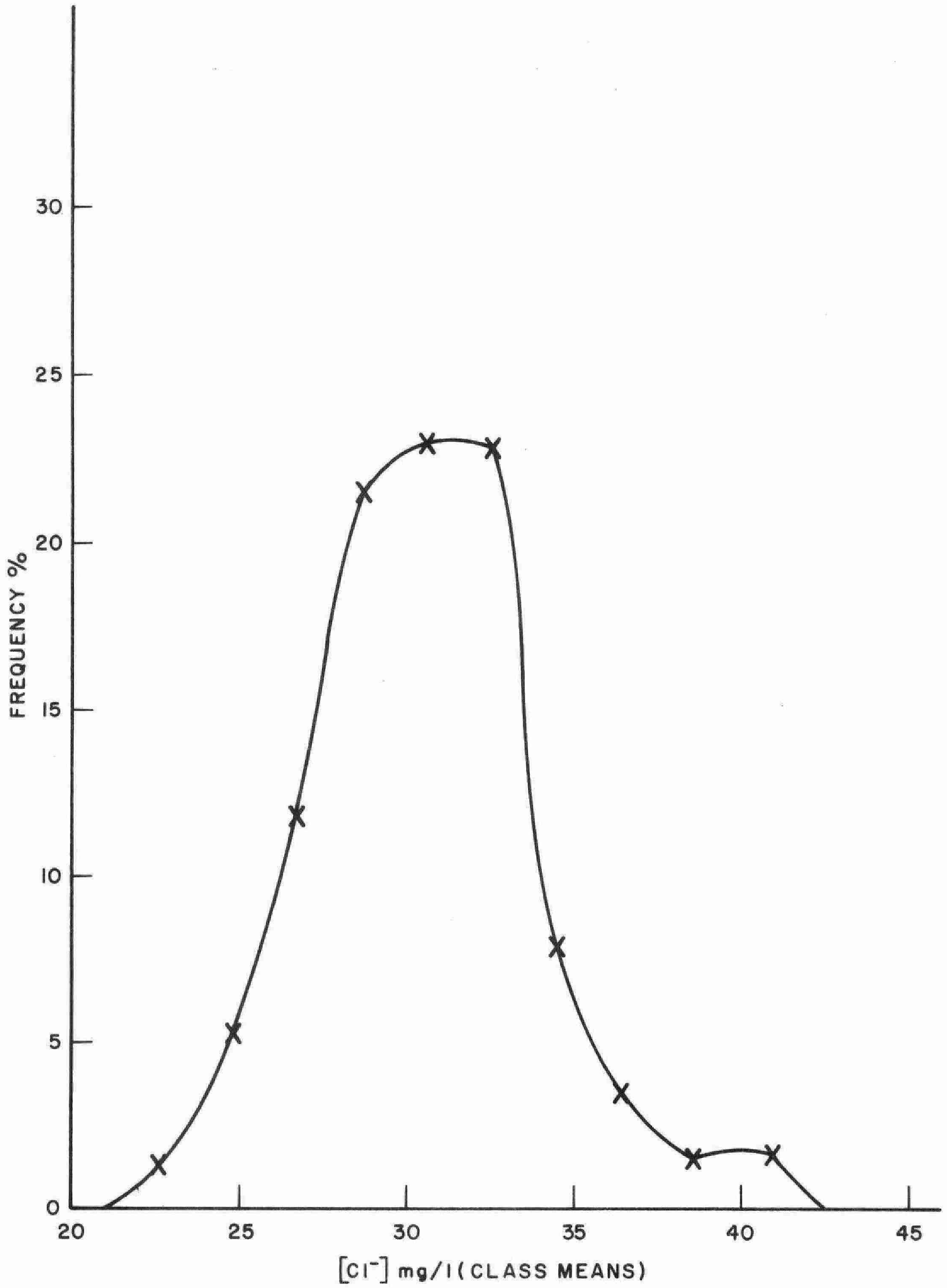

Fig. 4 Trend removed chloride, April 1971, frequency distribution

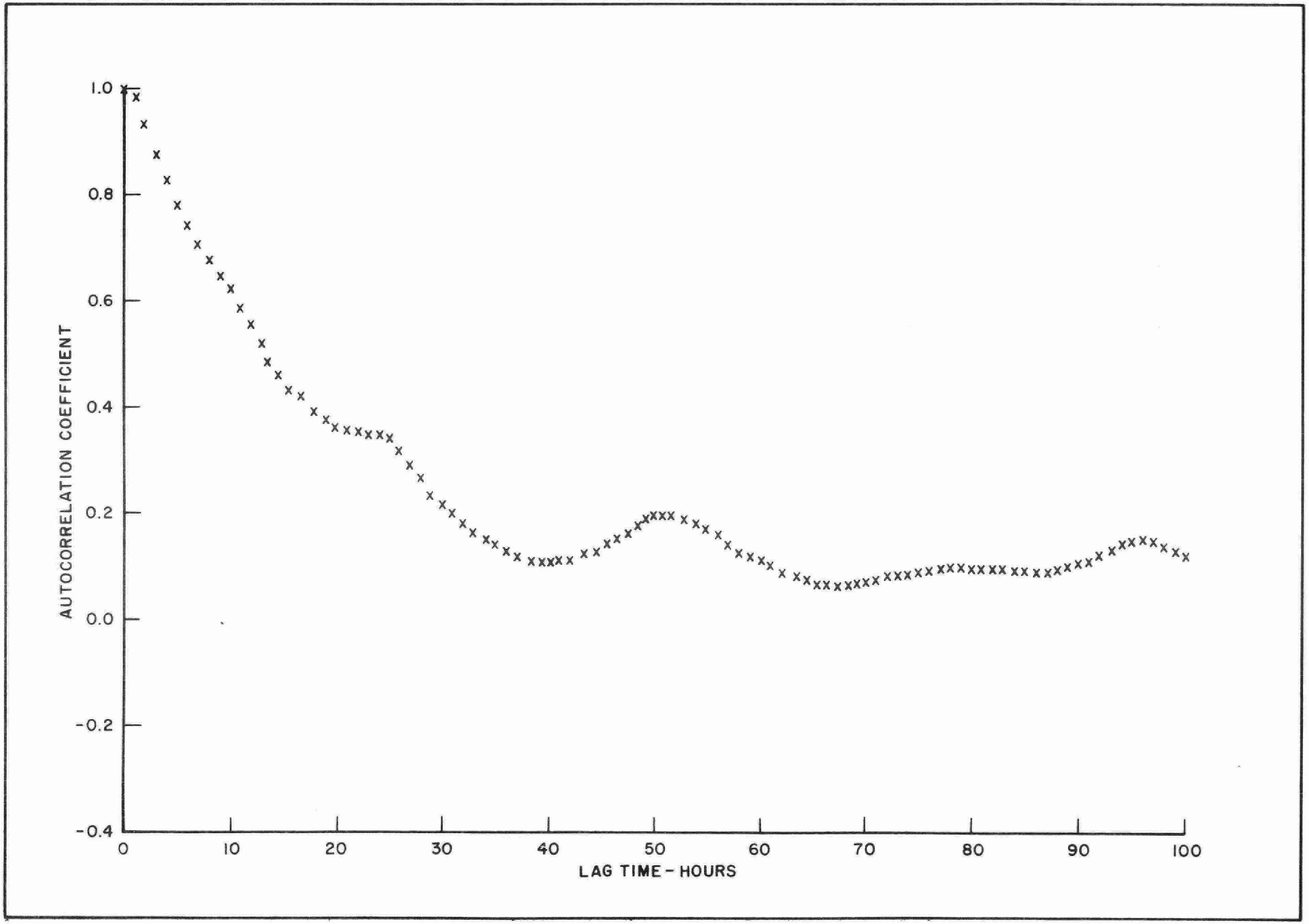

Fig. 5 Autocorrelation coefficients for June 1971 chloride

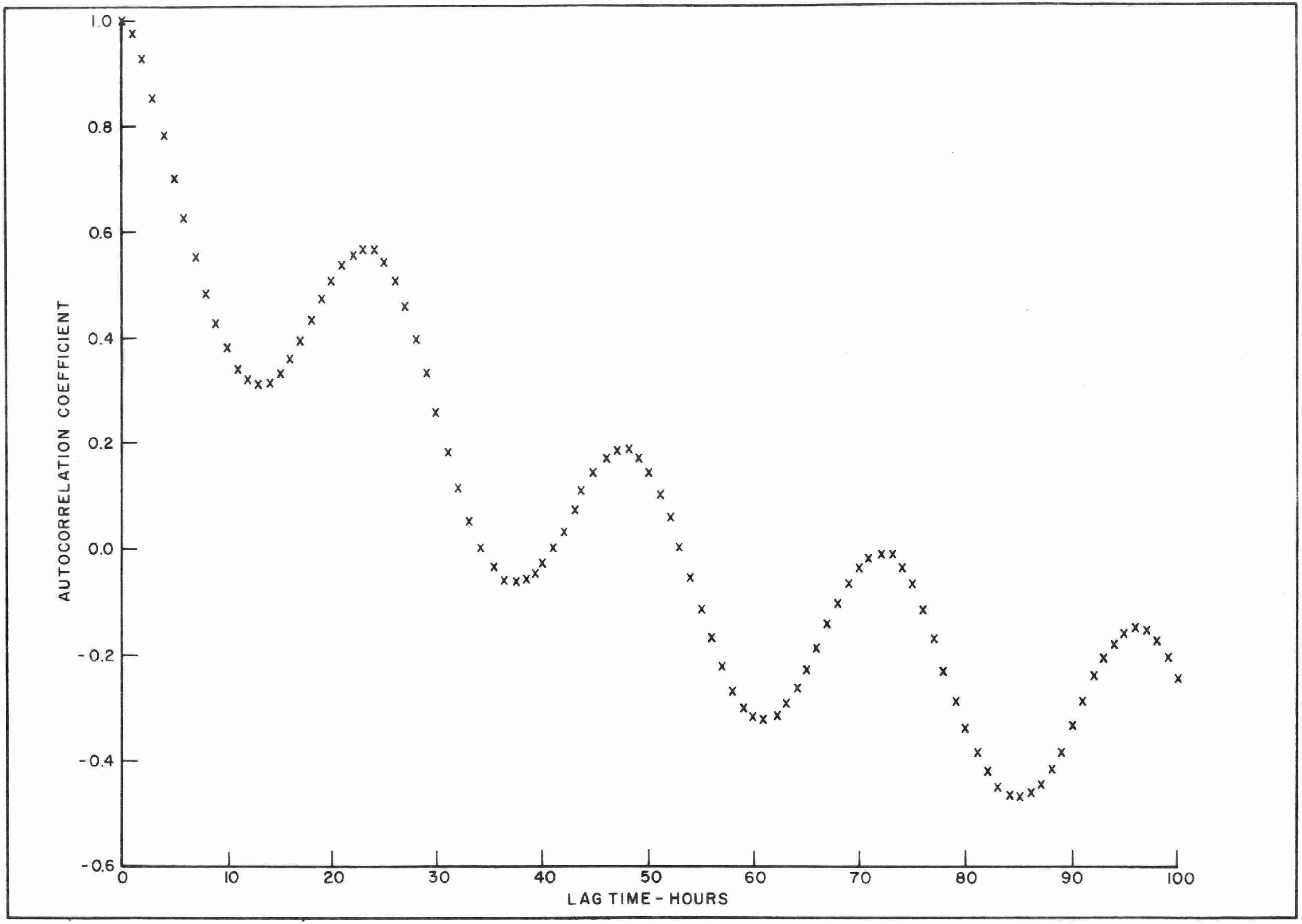

Fig. 6 Autocorrelation coefficients for June 1971 temperature

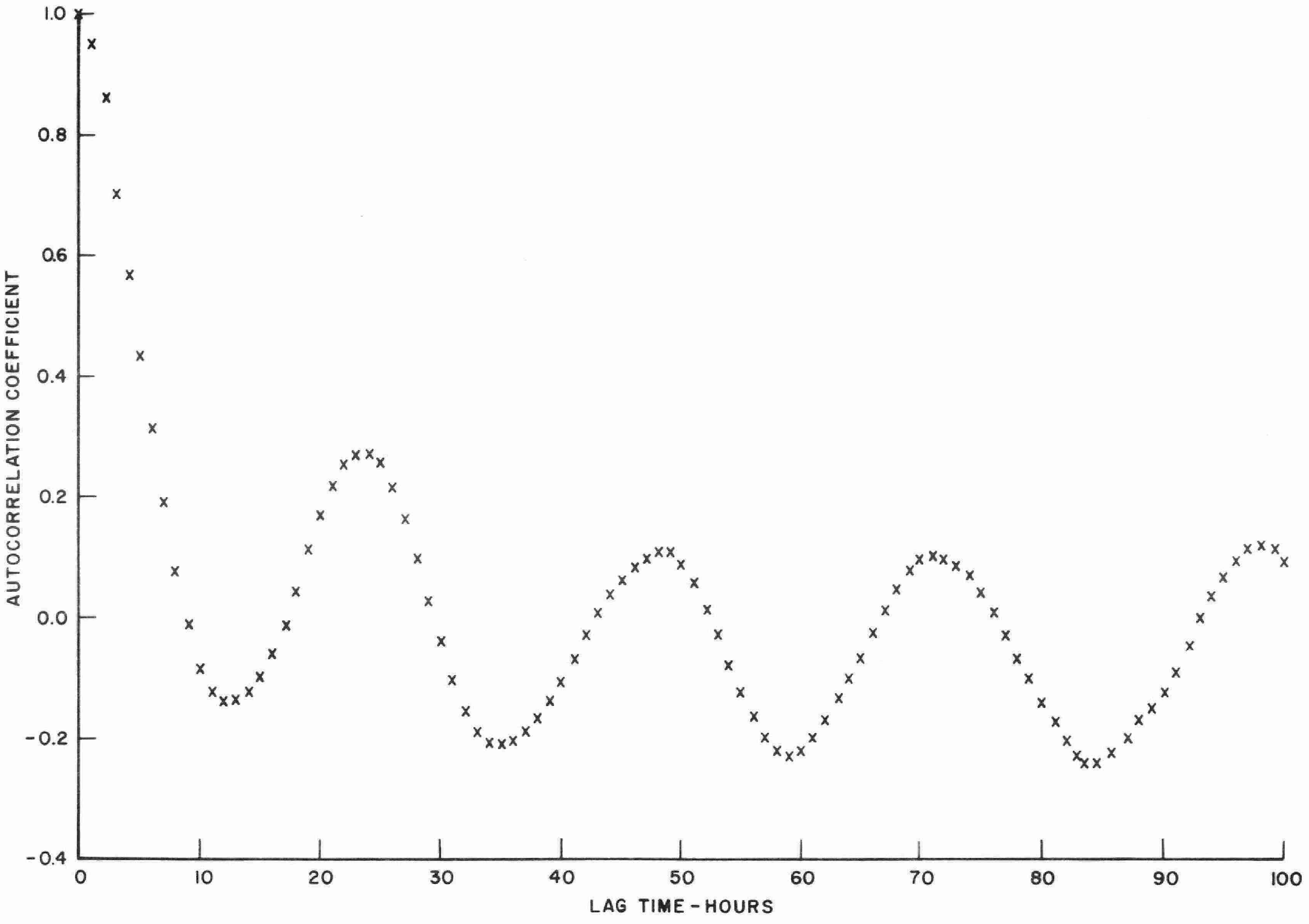

Fig. 7 Autocorrelation coefficients for June 1971 dissolved oxygen

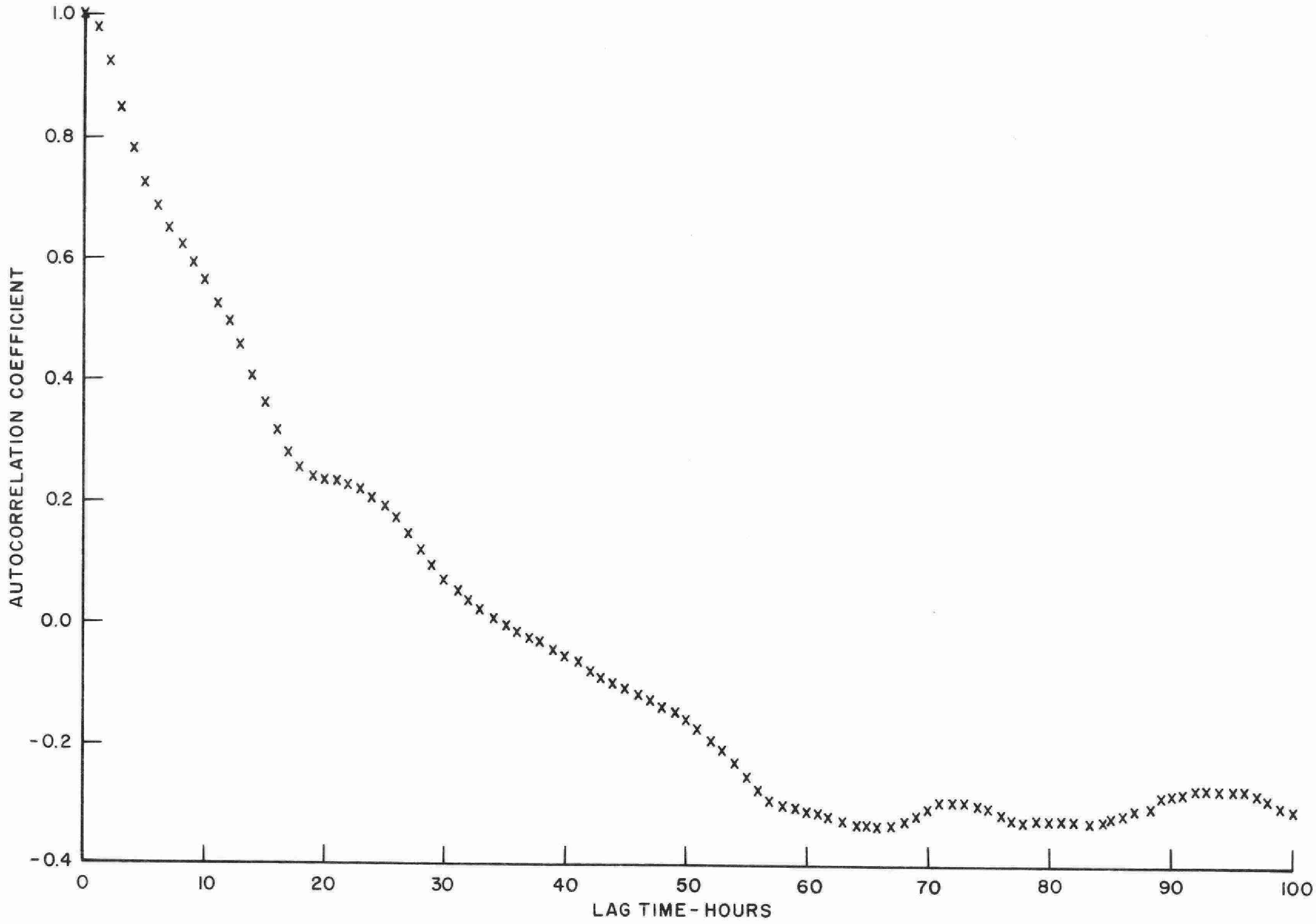

Fig. 8 Autocorrelation coefficients for June 1971 conductivity

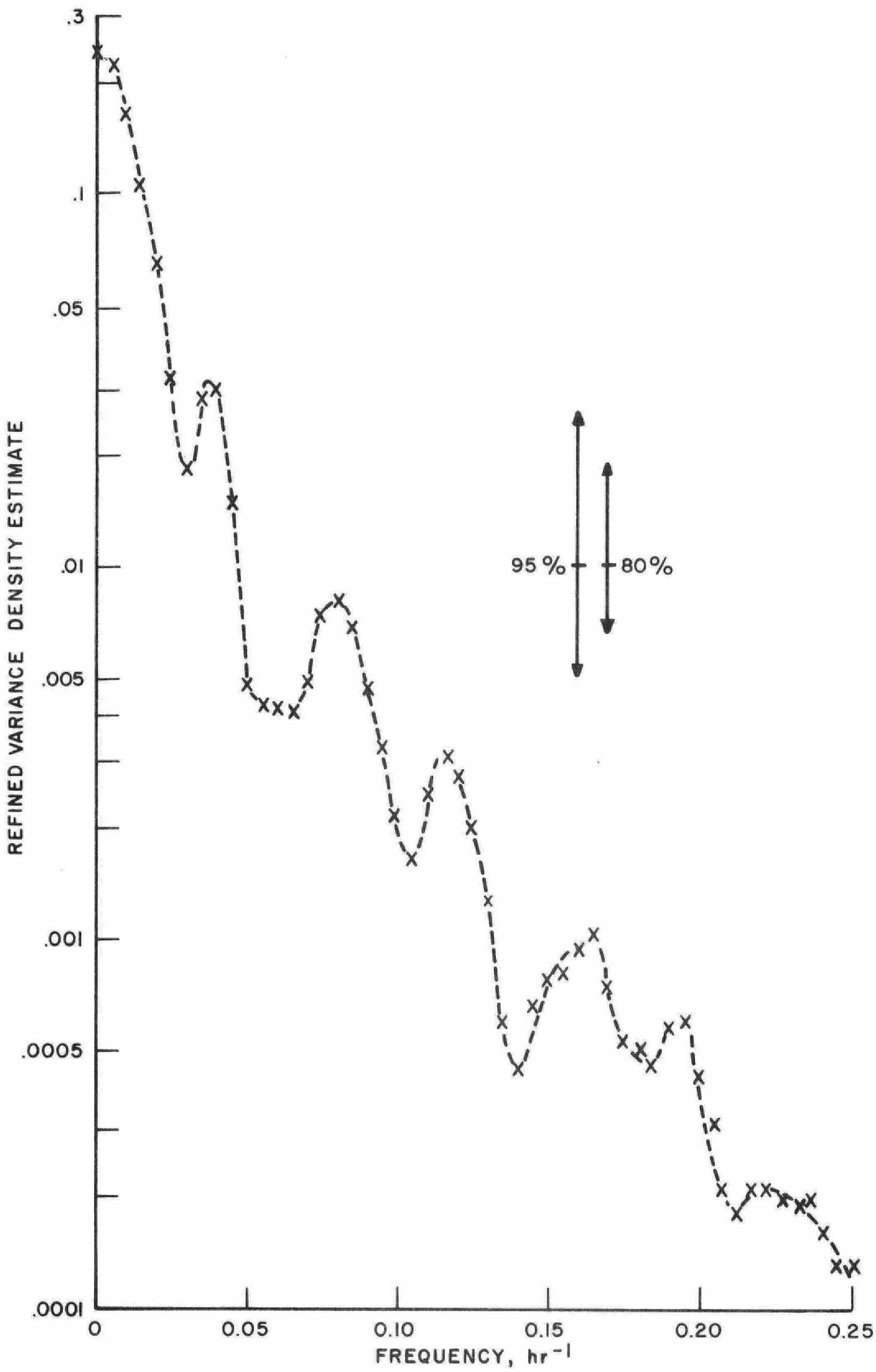

Fig. 9 Variance density spectra for June 1971 chloride

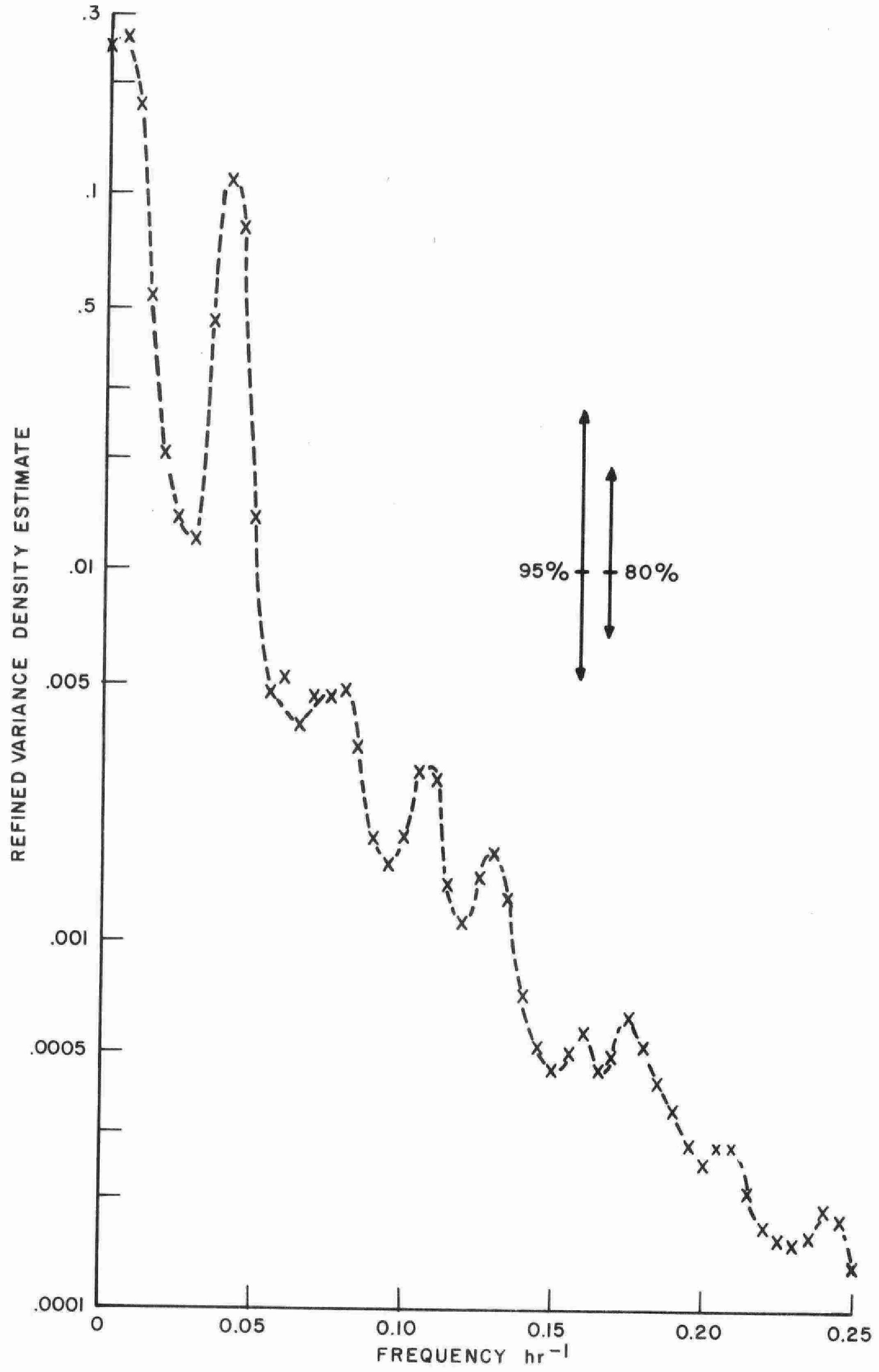

Fig. 10 Variance density spectra for June 1971 temperature

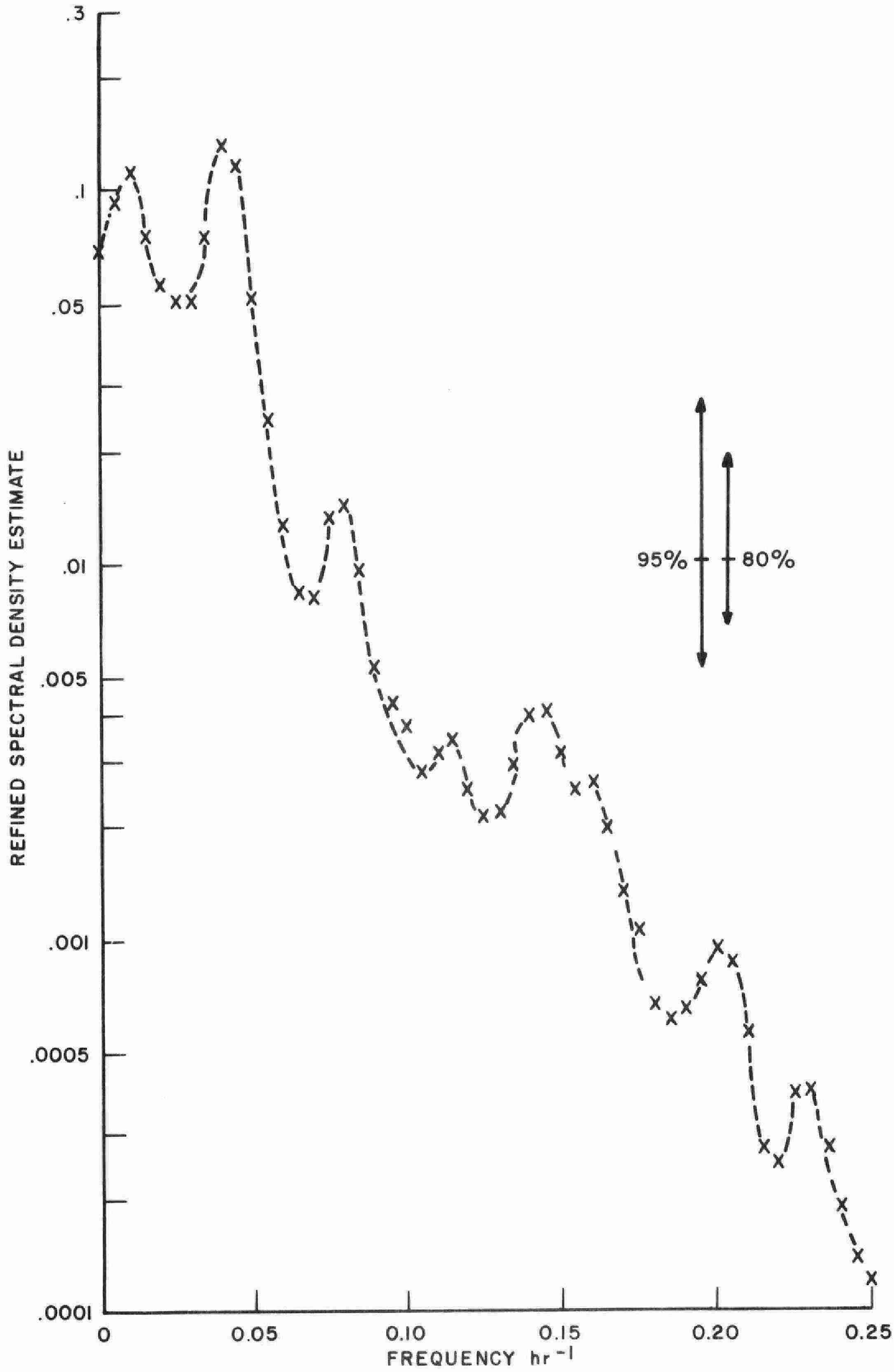

Fig. 11 Variance density spectra for June 1971 dissolved oxygen

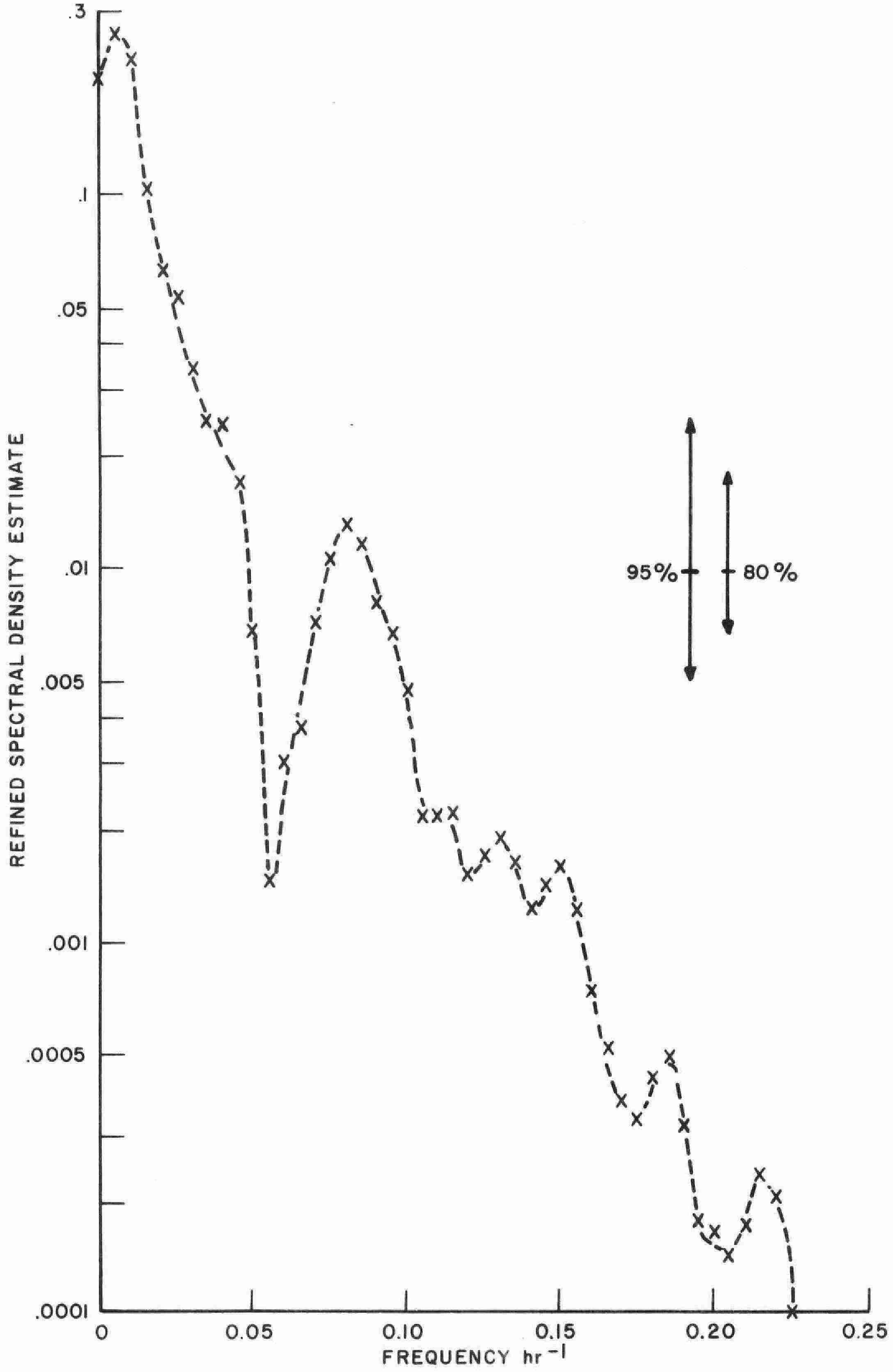

Fig. 12 Variance density spectra for June 1971 conductivity

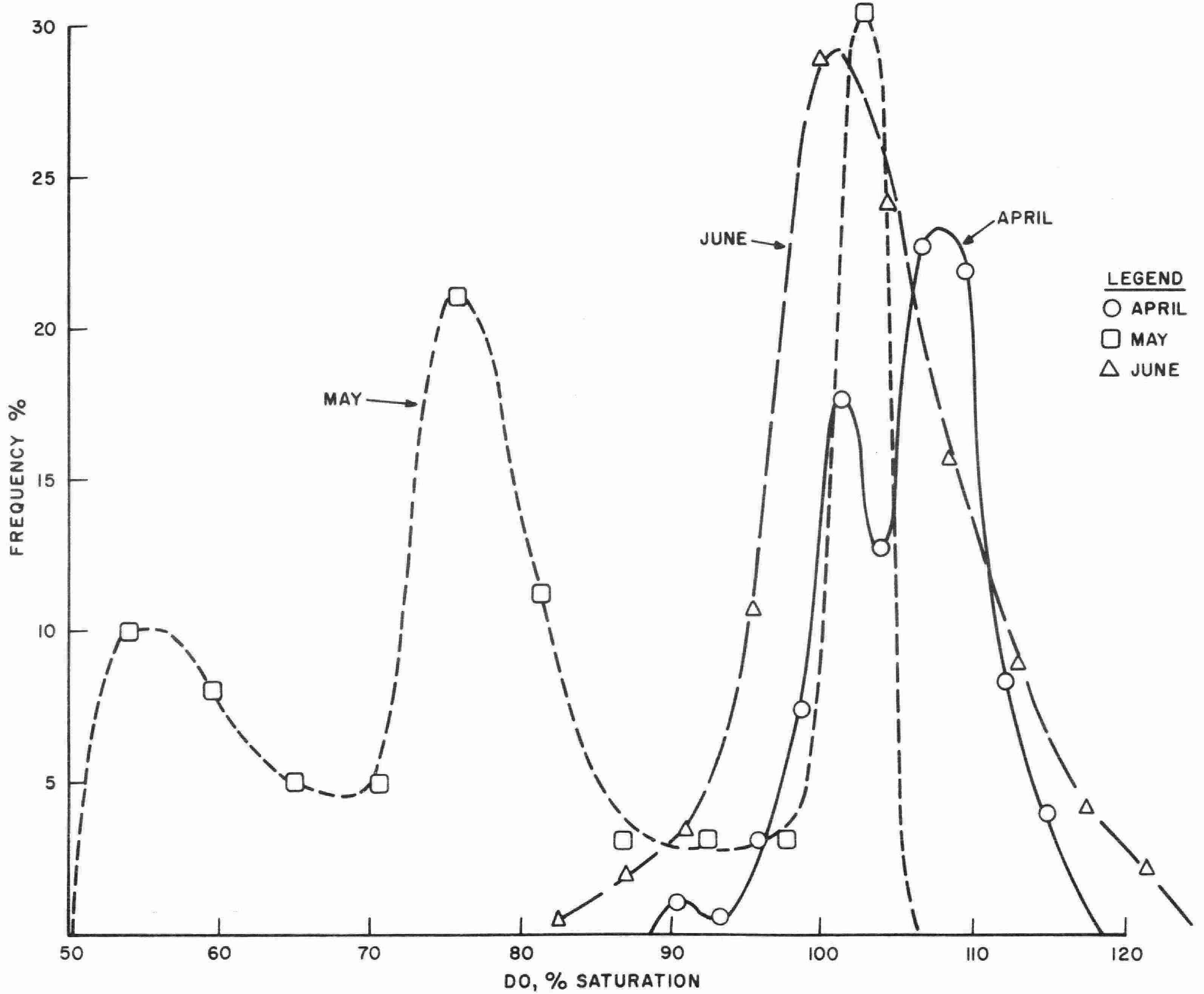

Fig. 13 Dissolved oxygen frequency distributions for April-June 1971

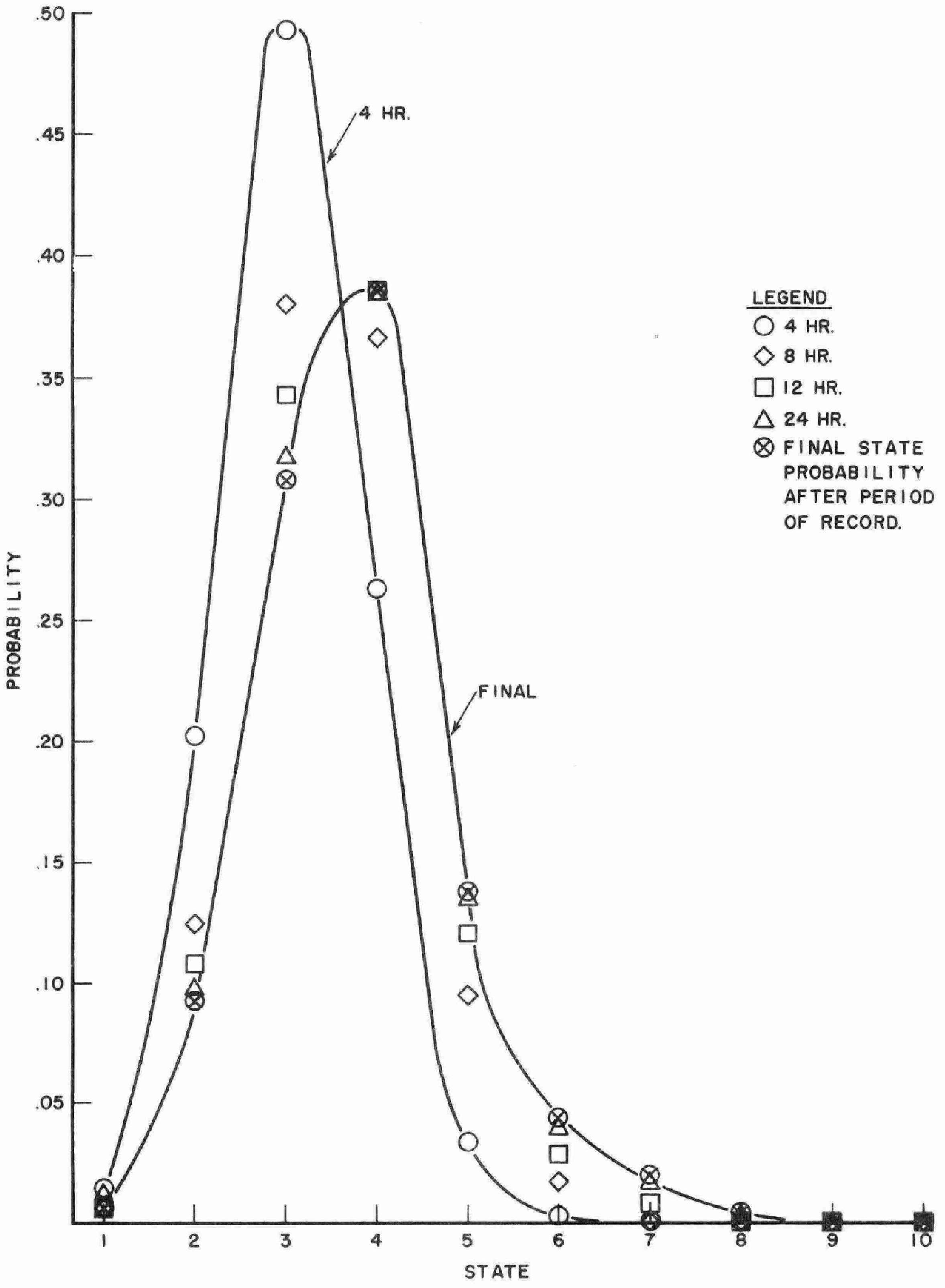

Fig. 14 State probability distributions for chloride, June 1971

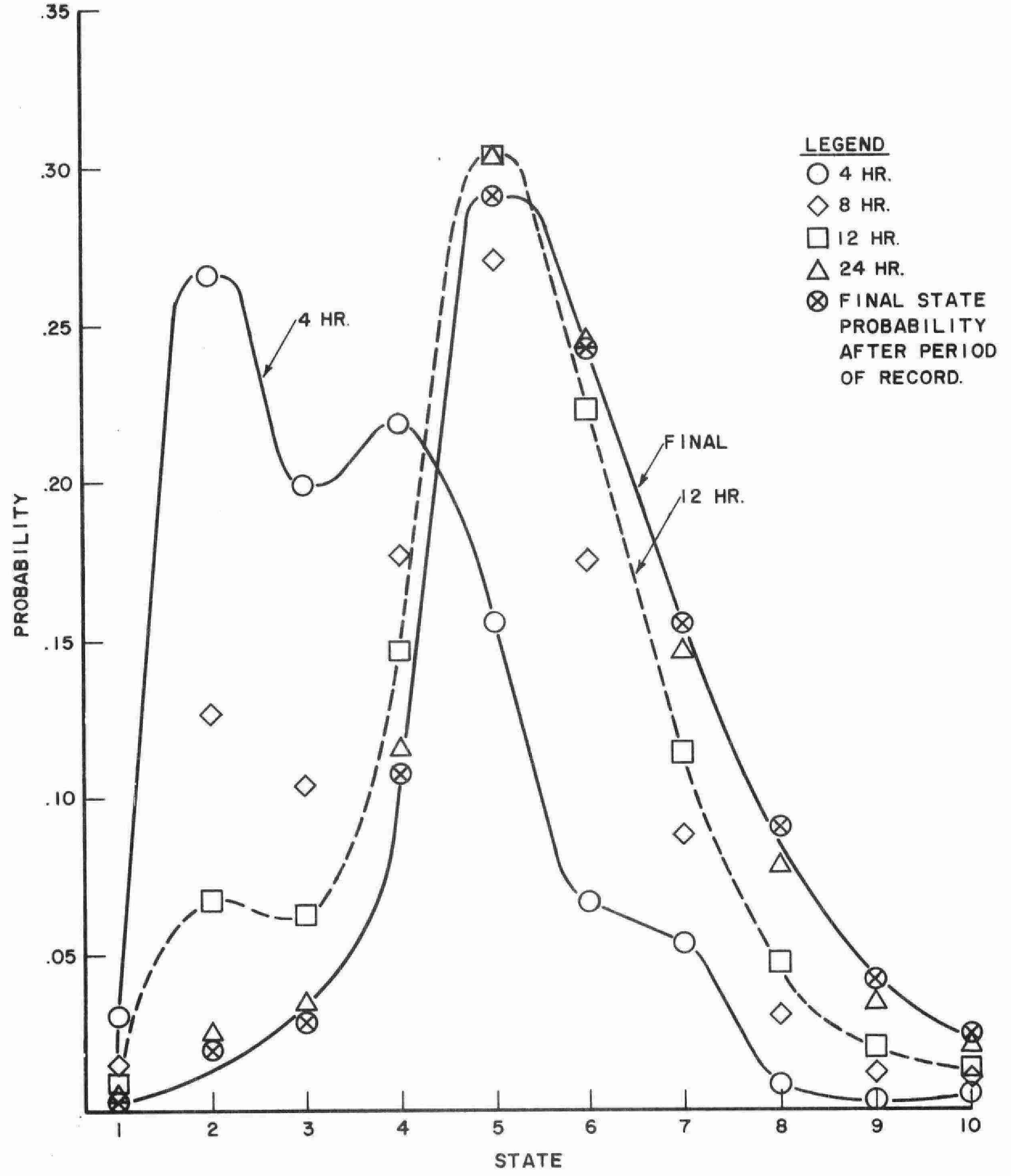

Fig. 15 State probability distributions for dissolved oxygen, June 1971

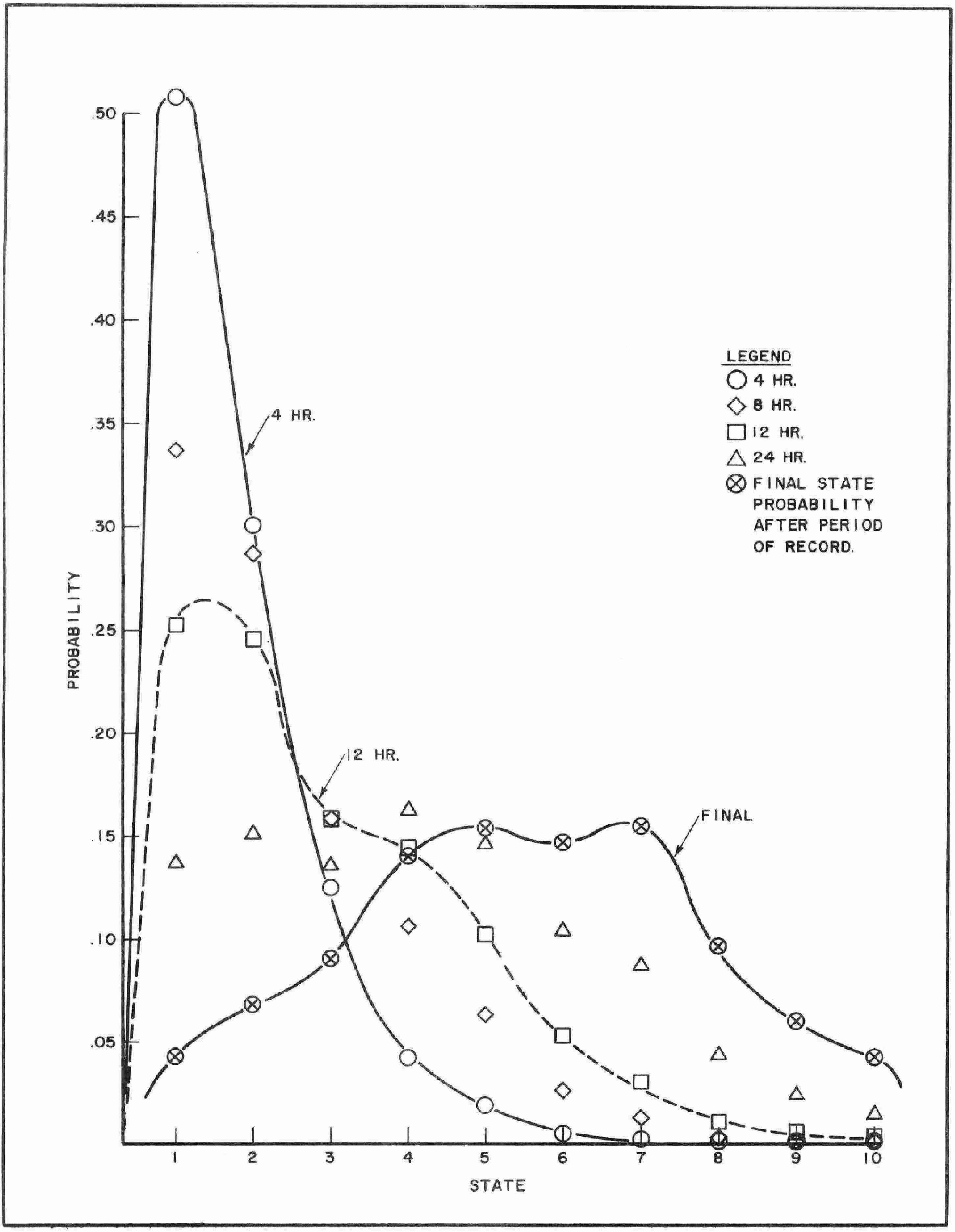

Fig. 16 State probability distributions for conductivity, June 1971

APPENDIX 1

DESCRIPTION OF THE ROBOT MONITOR AND PROBLEMS ENCOUNTERED DURING ITS OPERATION

A1.1 Description of Installation

The "robot monitor" unit is located in the Corunna waste treatment plant. The intake and pumping arrangement, as of August 1971, is shown in Figure A1.1.

The intake line is 30 meters of 3 cm diameter pvc hose, connected to 4 m of 10 cm diameter perforated pvc pipe directed downstream, and anchored to a steel tripod at a depth of 3.5 m. The piece of large tubing served to avoid uptake of weeds, sand or other debris on the river bottom. Concrete blocks were also attached along the hose at approximately 2m intervals, to keep the hose on the river bottom. This helps to prevent dragging of the line due to river currents resulting from passing ships or ice floes. A pump of 1 1/2 h.p. is used to provide a constant flow of 8-12 liters/min required for each cell. This pump, with an electric heater to maintain an adequate indoor temperature, is housed in a wooden hut sitting on a culvert. As this unit is entirely in the water, it is subject to action of currents and waves, as well as ice in the winter, and a more permanent concrete housing has been considered.

A 2.5 cm steel pipe about 4 m long connects the pump to a 2.5 cm plastic hose about 80 m long, buried at a depth of 0.5 - 1 m in porous soil, leading to the robot monitor.

Although freezing problems might be expected in the winter, difficulties were encountered only once (February 1971) in the four years the installation has been in use.

Al.2 Problems Associated with Installation

(1) Weeds

On July 29, 1971, the original intake line, which was about 10 m long at a depth of 1.5 m, was observed to be in a concentration of weed growth. The growth was worse on August 16, when the line was extended past the weeds into an area of clear water. However, on August 23, the flow had dropped from 20 to 6 l/min due to weeds plugging the intake line. At this time, the 4 m length of perforated 10 cm diameter pvc tubing was added, and flow conditions became normal.

(2) Ice

Action of ice in February and March 1972 resulted in the intake line being broken and carried away. Sand was taken in, and at one point the DO and pH cells were plugged. Before the extent of damage was known, attempts were made to keep the system in operation by backflushing once or twice a week. On May 29, the line was replaced, and proper flow conditions were again achieved.

A1.3 Problems Associated with the Instrument

(1) Digitizer

The digitizer was newly installed on April 14, 1971. Problems which have been encountered are as follows:

(a) At times the alignment of the punch head was incorrect. Frequently, the punch holes were at the edge of the tape. Because of this, a replacement punch head was used in the period May-July 1971, while the original was repaired. Afterward, occasional adjustment was required.

(b) The timer controlling the take-up motor is not reliable. This normally allows the take-up motor to run for a set interval following punching of a set of data, in order to wind up the punched tape. Failures of this unit have resulted in continuous operation of the take-up motor (Sept. 26/71) and failure of the take-up motor to operate (June 24, 1972). In both cases, the timer was replaced. Presently, this problem is being discussed with Schneider Instrument Co.

(2) Reliability of Probes

No problems were encountered with the temperature and conductivity probes, with the changes at calibration being within the expected precisions.

The DO probe presented considerable problems. As a result of silver chloride build-up on the silver anodes, which flaked off from time to time during operation, as well as slime deposition on the membrane, frequent reconditioning and recalibration was required. Comparison of digitizer reading with results obtained by Winkler titration indicated errors of ± 10 percent or more in the period April 1971-January 1972. Large negative errors were consistently found in February-March 1972, but this occurred when the intake was malfunctioning. In order to improve the reliability, the calibration of the probe should be checked weekly against the Winkler results, as was done after March 3, 1972. Weekly reports by S. Bechard before this date only give panel meter readings; it is suggested that digitizer readings, as well as chemical test results be reported weekly for all parameters.

The chloride probe was reliable over most of the time period, with only small adjustments being needed on recalibration. A notable exception was in the May 4-May 27, 1971 period, in which the ratio digitizer: analyzer reading was 3 to 3.5, rather than the expected 2. It was stated that calibration was performed but no reason for the observed differences was given.

Reliability of the pH probe was comparable to the chloride probe. At times, crystallization of KCl in the reference electrode resulted in unreliable results such as in June 1971. A serious error also was noted on November 22, 1971, when the pH 8 buffer read 9.10. At other times, recalibration adjustments amounted to ± 0.1 - 0.2 pH units.

A1.4 Accuracy of Equipment

Estimated accuracies, as claimed by Schneider Instrument Company, are as follows: temperature ± 1°F; conductivity, DO and pH, ± 1%; chloride, ± 5%. The ranges of measurement were: conductivity, 0-1200 and 0-2400 micromhos; DO, 0-12 and 0-24 mg/l; chloride, 0-600 mg/l.

A1.5 Maintenance of Equipment

Periodic visits to Corunna have been performed by J. Gabay, and the results of these visits recorded in the log book. During these visits, all sensors and other equipment were checked and required service performed as suggested in the instruction manual. Meter readings of the various parameters were compared with digitizer readings, and the results recorded. Where necessary, recalibrations were performed. From September 17, 1971, weekly checks were made by S. Bechard.

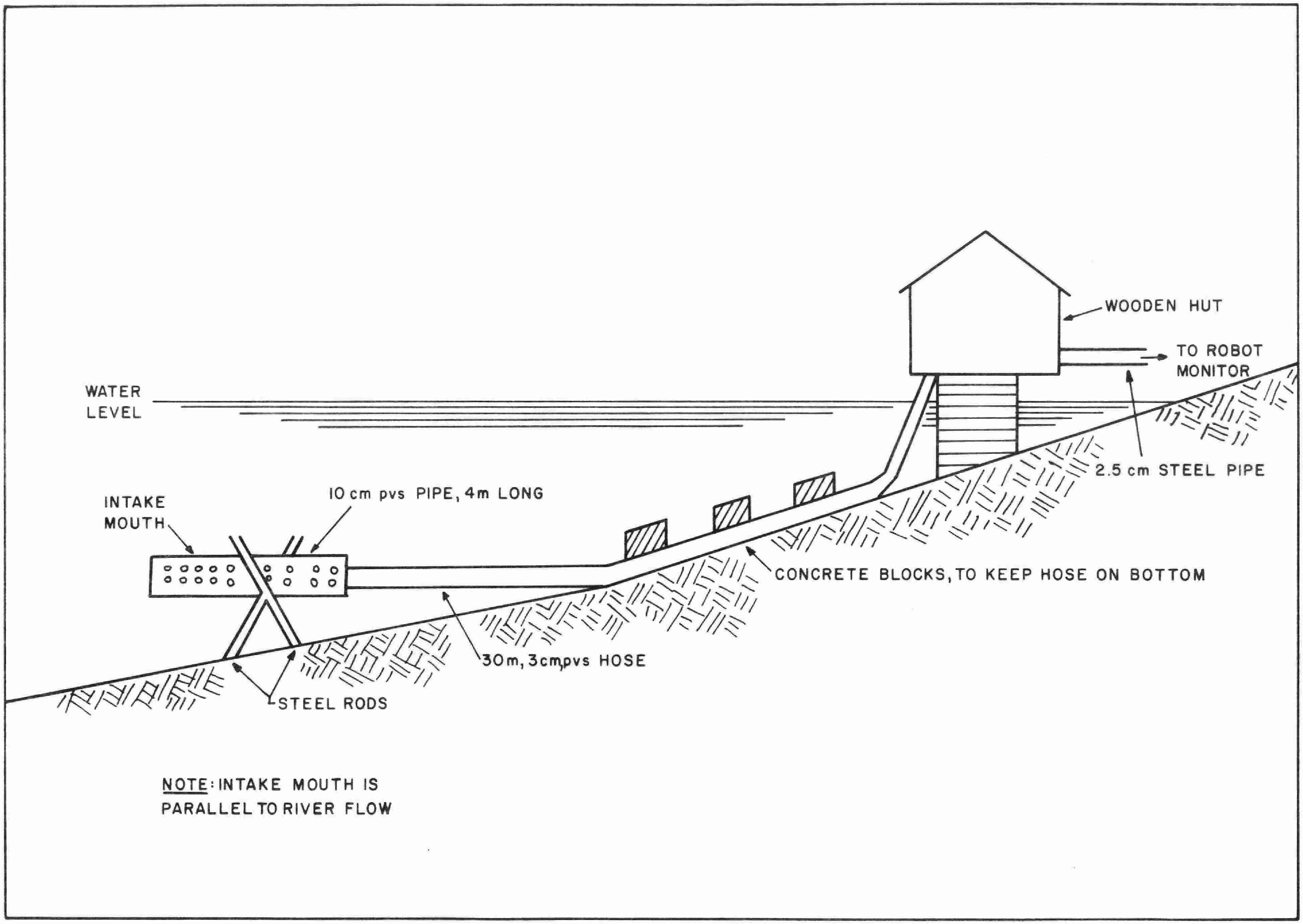

Fig. A1.1 Schematic drawing of robot monitor intake August 1971

APPENDIX 2

COMMENT ON QUALITY OF VARIOUS DATA RECORDS

To facilitate discussion of the best data obtained from the Corunna robot monitor, comments on all data obtained have been included in Table A2.1. The best records were obtained for April-June 1971 and December 1971, and these are discussed in detail elsewhere, along with any data from several "fair" records, as applicable. In July 1971, calibration difficulties were experienced. In August and September 1971, the paper tape was punched off center, although the September tape was translated following re-punching. The November 1971 record was only fair due to calibration difficulties.

During 1972, very little, if any, good data were obtained. Problems with ice, which have been mentioned in Appendix 1, resulted in approximately three months' data being obtained from the stream bottom, where considerable mud was drawn in by the pump, clogging sensors several times. The data for February 1972, which were obtained prior to the start of ice difficulties,are believed to be good, except for dissolved oxygen, which was out of calibration. During the period May to July 1972, electronic difficulties of gradually increasing magnitude were experienced. Failure of a stepping relay in the digitizer during this time period resulted in records of alternating high and low readings, which could not be translated meaningfully. At the end of July 1972, the instrument was taken out of service for repairs.

TABLE A2.1

INDEX OF DATA TAPES OBTAINED AT CORUNNA

Location Code	Period of Operation	Text Reference	Comments
81901	Apr 15 - May 4/71	April 1971	good record, 2 hr. frequency
81903	May 20 - May 27/71	May 1971	good record, Cl^- out of calib.
81904	May 28 - June 23/71	June 1971	good record, except pH
81905	June 23 - July 28/71	July 1971	Cl^-, DO, pH no good
81906	July 29 - Aug 16/71	August 1971	tape punched off center
89107	Aug 17 - Sept 14/71	Sept 1971	fair record, no good periodicities
81908	Sept 30 - Oct 14/71	early Oct. 1971	poor record, high deviations
81909	Oct 14 - Oct 25/71	late Oct. 1971	poor record, high deviations
81910	Oct 25 - Nov 16/71	early Nov. 1971	pH and Cl^- out of calibration
81911	Nov 16 - Nov 26/71	late Nov. 1971	Temp and DO questionable. pH out of calibration
81912	Nov 26 - Dec 25/71	Dec 1971	good record
81913	Jan 14 - Jan 20/72	Jan 1972	fair record
81914	Jan 20 - Feb 27/72	Feb 1972	fair record, DO no good
" *	Mar 2 - Apr 7/72	Mar 1972	intake at bottom
81915	Apr 7 - Apr 17/72	Apr 1972	intake at bottom
81916	Apr 17 - May 30/72	May 1972	intake at bottom
81917	May 30 - June 25/72	June 1972	poor record, instrument needed overhaul
81918	June 29 - July 27/72	July 1972	as for June 1972

Note: The text reference is an approximate date reference used in referring to the various tapes in the text and other tables.

* This record was split into two portions due to its extreme length.

APPENDIX 3

STATISTICAL ANALYSIS OF FREQUENCY DISTRIBUTIONS

A comparison of several statistical methods of analysis of robot monitor data for normal, exponential and log normal distribution was performed. The following methods were selected.

1. Kolmogorov-Smirnov (continuous distribution)
2. Von Mises (equation for normal distribution)
3. Von Mises (discrete distribution)
4. χ^2 test (discrete distribution)

KOLMOGOROV-SMIRNOV TEST

The Kolmogorov-Smirnov test, which is designed for comparing two distribution functions, is the test which has previously been used (IBM, 1968). In this test, the data (expressed in terms of number of standard deviations from the mean) are sorted into ascending order. The maximum deviation D from the theoretical distribution is then obtained, and the test statistic Z is calculated as:

$$Z = D\sqrt{n}, \qquad [A1]$$

where n = number of readings.

To test the effect of sample size on the probability and to compare this test with the Von Mises test for normal distribution, a set of chloride data was sampled and tested for normal and exponential distribution. The results are given in Table A3.1.

The value of Z for exponential distribution is approximately inversely proportional to the square root of the sampling frequency. For normal distribution, the rate of decrease is less. At any rate, for very small numbers of readings, the probability of normal distribution also increases rapdily. This illustrates the hazard of attempting distribution calculations on small data samples.

VON MISES TEST (Equation for Normal Distribution)

Von Mises (1964) devised a test which is applicable to both discrete and continuous distributions. The observed distribution function S(x) is compared to the theoretical distribution function P(x), and ε is defined as:

$$\varepsilon(x) = n\,[S(x) - P(x)] \qquad [A2]$$

where n is the number of observations. As a measure of the deviation, a parameter ω^2 was defined as:

$$\omega^2 = \lambda(x)\ \varepsilon^2(x)dx \qquad [A3]$$

where λ is a weight function, supposedly evaluated to make the expected value of ω^2 equal to 1. In the case of the normal distribution, it can be shown (Von Mises, 1964) that:

$$\omega^2 = 2\sqrt{\pi}\sum_{i=1}^{n}\left(\phi(x_i) + x_i\,[\Phi(x_i) - (2i-1)/2n]\right) - n \qquad [A4]$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-x^2}{2}}\,dx$, the equation for the normal distribution

and $\phi(x) = \Phi'(x)$

The standard deviation of ω^2 is given by:

$$\sigma = \sqrt{0.63 - 0.12/n} \qquad [A5]$$

Finally, the probability of finding the normal distribution is calculated from the normal probability curve for the deviation of ω^2 (in terms of σ) from its expected value.

The Von Mises test for normal distribution was checked for several samples. It was found, however, that $E(\omega^2) = 0$, rather than 1, as stated by Von Mises. This was verified by taking a data set of 100 readings of normal distribution, taken from the normal probability table. For this data $\omega^2 = 0.0078$. Some typical probabilities are lower than those obtained by the Kolmogorov-Smirnov test, and also increase with decreasing sample size. As well, this method is not suited to calculating probabilities of exponential distribution. Therefore, its use was not further considered.

VON MISES TEST FOR A DISCRETE DISTRIBUTION

The data are divided into a given number of classes, each class representing an equal fraction of the difference between the minimum and maximum reading. The number of observations n_i within each class i is determined. ω^2 is then calculated from the following equation.

$$\omega^2 = \sum_{i=1}^{k} \frac{(nQ_i - N_i)^2}{n} \cdot \frac{\sqrt{\pi}}{\sigma} \qquad [A6]$$

where $N_i = \sum_{j=1}^{i} n_j$ = the number of observations in all classes below and including class i

Q_i = the normal cumulative distribution function for the mean of class i

n = the total number of readings

σ = the standard deviation of all readings

k = the total number of classes.

The standard deviation of ω^2 is calculated from equation [A5] as in the continuous case, and a probability of normal distribution can be estimated from the normal curve. If desired, the probability of exponential distribution may be estimated by fitting Q_i from an exponential curve.

Edited chloride data from December 1971 were subjected to this test. The results are given in Table A3.2. The value of ω^2 is directly proportional to the same size n, as is evident from the equation. Also, ω^2 increases as the number of classes into which the sample is divided is increased. These drawbacks produce severe limitations on the validity of using the Von Mises test for discrete distributions on the present data.

CHI-SQUARE TEST (Discrete Distribution)

A frequency distribution is obtained as discussed for the Von Mises test. A normal distribution is fitted, and expected frequencies obtained. The value of χ^2 is then obtained from the equation (Snedecor and Cochran, 1967):

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - F_i)^2}{F_i} \qquad [A7]$$

where k = number of classes

f_i = observed frequency in class i

F_i = calculated frequency in class i.

The probability is calculated from the χ^2 distribution with (n-3) degrees of freedom. Fitting the exponential distribution presents the drawback that terms in χ^2 become infinite for data more than one standard deviation below the mean. This difficulty was circumvented by replacing F_i in the denominator of the above equation by f_i when F_i was close to or equal to zero.

Results of this test on edited chloride data from December 1971 are presented in Table A3.2. As in the Von Mises test, the value of χ^2 is directly proportional to the number of observations; the effect on the probability is even worse than that observed with the Von Mises test, and allows the probability to assume values in almost the complete range of 0 to 1. The value of χ^2 also increases with increasing number of classes, but this change is compensated for by the increased number of degrees of freedom.

As a result of these drawbacks, it was decided to continue using the Kolmogorov-Smirnov test for estimating distribution probabilities.

SUMMARY

The performance of several statistical tests for evaluating probabilities of normal, exponential and log-normal distribution was evaluated. All methods suffer from the drawback that the observed probabilities are dependent upon the sample size, but this dependence is worst for the discrete distribution tests. The following additional drawbacks are noted:

1. In the Von Mises tests (both continuous and discrete), the expected value of ω^2 is 0, not 1 as claimed by Von Mises.
2. The Von Mises continuous test is unsuitable for exponential distribution.
3. Probabilities for the discrete Von Mises test decrease as the number of classes in the frequency table is increased.

As the Kolmogorov-Smirnov test is least affected by these drawbacks, it will continue to be used in estimating the distribution probabilities.

TABLE A3.1

KOLMOGOROV-SMIRNOV AND VON MISES TEST RESULTS ON SAMPLED DATA

Data used are edited Cl^- from January 1972. (144 readings).

Sampling	Exponential		Normal		Von Mises	
Frequency	Z	Prob	Z	Prob	ω^2	Prob $[E(\omega^2)=0]$
1	2.083	.0003	.850	.466	1.041	.189
2	1.482	.025	.593	.874	.456	.565
3	1.167	.131	.561	.911	.402	.612
5	.929	.355	.433	.992	.245	.757
6	.817	.518	.420	.995	.229	.773
10	.668	.763	.369	.999	.161	.838
12	.577	.893	.336	.9999	.132	.867

TABLE A3.2

DISCRETE NORMAL DISTRIBUTION TESTS

Data used are edited Cl^- from December 1971 (689 readings)

No. of Classes	Method of Expressing Frequencies	Von Mises		Chi-Square	
		ω^2	Prob $[E(\omega^2)=0]$	χ^2	Prob
20	No.observations	1.114	0.159	46.24	0.0002
20	percent	0.162	0.838	6.71	0.987
10	No. observations	0.294	0.710	19.91	0.006
10	percent	0.0427	0.957	2.89	0.895

Note: The Kolmogorov-Smirnov probability for normal distribution of the above data is 0.025.

REFERENCES

INTERNATIONAL BUSINESS MACHINES, 1968. System 360 scientific subroutine package, version III, programmer's manual. pp 63-68, 433-438.

SNEDECOR, G.W. and W.G. COCHRAN, 1967. Statistical methods, 6th ed. Iowa State University Press, Ames, Ia. 593 pp.

VON MISES, R., 1964. Mathematical theory of probability and statistics. Academic Press, New York.

APPENDIX 4

OPERATIONS AND CALIBRATIONS LOG

1. Operations Log

Date	Service
April 14/71	Digitizer installed; DO probe replaced.
May 19/71	Digitizer skipping punches and off center; alignment corrected.
May 27/71	Digitizer punch head replaced.
June 23/71	DO probe replaced. pH and Cl^- reference electrodes filled with KCl solution.
July 16/71	DO probe replaced.
July 28/71	Original punch head replaced. Intake in weed growth.
Aug. 16/71	Intake extended, anchored with cement blocks. pH and Cl^- reference electrodes cleaned, filled.
Aug. 25/71	Intake extended, 10 cm pvc pipe of length 3.5 m installed.
Sept. 15/71	pH and Cl^- reference electrode cleaned, filled. DO probe replaced.
Sept. 26/71	Digitizer malfunctions: improper spacing of punched tape, takeup motor running continuously. Timer replaced.
Sept. 30/71	Punch unit replaced, spacing adjusted. Conductivity sensor repaired.
Nov. 22/71	DO probe replaced. pH and Cl- reference electrodes cleaned and filled.
Jan. 13/72	Adjusted punch unit, serviced digitizer. Fixed open ground on pH probe DO probe replaced.
Mar. 6/72	Intake line removed by ice. Frequent backflushing of intake performed.

April 17/72	Attempt to repair intake unsuccessful due to ice movement in river.
May 29/72	Stepping relay cleaned; intake line repaired. pH and Cl^- reference electrodes cleaned and filled.
June 29/72	Timer of takeup motor replaced. Stepping relay cleaned.
Aug. 4/72	Monitor taken out of service for repairs.

2. Sample Calibration Log. (May 27/71)

Note: Routine calibration was performed on each of the above dates.

Parameter	Analyzer Reading	Digitizer Reading	Measurements Taken
Temperature	51°F	512	52.7°F
Conductivity	355 µmho	354	350 µmho; calibrated on standards 140, 680, 970 µmho
DO	12.75 mg/l	599 (=11.98 mg/l)	12.8 mg/l
Cl^-	30 mg/l	106 (=53 mg/l)	50 mg/l
pH	9.6	838	Calibrated on buffers pH=7 and 4